

**Mastergradsoppgave i Molekylær biovitenskap**  
*Studieretning biokjemi*

**Ahp A-domener i cyanobakteriell peptid  
syntetase**

*Kari Elisabeth Amlie*



**Institutt for molekylær biovitenskap**  
Det matematisk-naturvitenskapelige fakultet

UNIVERSITETET I OSLO  
Juni 2007

# Forord

Denne oppgaven er utført i tidsrommet februar 2006 til juni 2007, under veiledning av professor Tom Kristensen og professor Kjetill Jakobsen ved Universitetet i Oslo.

Arbeidet med masteroppgaven har vært veldig spennende og lærerikt, og jeg vil gjerne takk Tom Kristensen og Kjetill Jakobsen for at de har gitt meg muligheten til å være en del av dette prosjektet. De er inspirerende forskere, med mye kunnskap og gode råd, og det vært utrolig lærerikt å være en del av deres gruppe.

I tillegg til professorene vil jeg takke stipendiat Ave Tooming-Klunderud for all hjelp og gode råd. Takk til Randi Ose for god støtte i det praktiske arbeidet, og takk til medstudent Tom Inge Sønju for godt samarbeid og hyggelige dager i laboratoriet.

Takk også til familie og venner for at dere tror på meg, støtter og hjelper meg så jeg når mine mål.

Juni 2007

Kari Elisabeth Amlie

---

## Sammendrag

Cyanobakteriestammer fra slektene *Anabaena*, *Microcystis*, *Planktothrix* og *Nostoc* er vidt utbredt i elver og sjøer over hele jorden. Disse og andre cyanobakterier bruker store multifunksjonelle enzymer, ikke-ribosomale peptid syntetaser (NRPS), til å produsere peptider med stor variasjon i struktur. Ved ikke-ribosomal peptidsyntese aktiverer adenyleringsdomenet (A-domenet) i hver modul hver sin aminosyre. Cyanopeptolin er en av de vanligste klassene av peptider som produseres av *Microcystis*, *Anabaena* og *Planktothrix*. Cyanopeptoliner er sykliske depsipeptider som inneholder aminosyrederivatet 3-amino-6-hydroksypiperidon, Ahp.

I denne oppgaven har cyanopeptolin syntetasegener blitt identifisert hos 14 stammer av cyanobakterier ved bruk av konsensus PCR primere som skiller mellom Ahp-aktiverende A-domener, og A-domener som aktiverer andre domener. Ut fra sekvensanalyse og fylogenetiske undersøkelser vises det at sekvensene fra A-domenene danner grupper ut fra hvilket substrat de aktiverer, og Ahp sekvensene danner en egen gruppe. Det er vist at rekombinasjon foregår mellom stammer av *Microcystis*.

Ahp blir innført i den voksende peptidkjeden fra en ukjent forløper, mest sannsynlig glutamat, glutamin, arginin, ornitin, eller glutaminsyresemialdehyd. Denne forløperen blir aktivert som en aminoacyladenylat av Ahp A-domenet i peptidsyntetasen, og jeg har gjort forsøk på å identifisere hvilken aminosyre det kan være som aktiverer Ahp ved proteinekspresjon og ATP/PPi utbytting.

# Forkortelser

A-domene	adenyleringsdomene
Adp	cyanopeptolin syntetase operon hos <i>Anabaena</i>
Ahp	3-amino-6-hydroksypiperidon
amp	Ampillin
ATP	adenosintrifosfat
BLAST	Basic Local Alignment Tool
bp	Basepar
BSA	Bovin serumalbumin
C-domenet	Kondenseringsdomene
Chl	Kloramfenikol
Cmp	telling pr. minutt
Da	Dalton
DMSO	dimetylsulfoksid
DNA	deoksyribonukleinsyre
dNTP	deoksynukleosidtrifosfat
DTT	Ditiotreitol
<i>E. coli</i>	<i>Escherichia coli</i>
EDTA	etylendiamintetraeddiksyre
EtBr	Etidiumbromid
IPTG	isopropyl $\beta$ -D-1-tiogalaktopyranosid
LB	Luria-Bertani-medium
Men	cyanopeptolin syntetase operon hos <i>Microcystis</i>
MEGA	Molecular Evolutionary Genetics Analysis
mRNA	messenger RNA
MW	molekylær vekt
NCBI	National Center for Biotechnology Information
NIES	Institute for Environmental Studies, Japan
NIVA	Norsk institutt for vannforskning
NJ	Neighbor Joining
NRPS	Ikkeribosomal peptidsyntetase (Non ribosomal peptide synthetase)
Oci	cyanopeptolin syntetase operon hos <i>Planktothrix</i>
O.D.	optisk tetthet
ORF	åpen leseramme
PAGE	Polyakrylamidgelelektroforese
PCC	Pasteurinstituttet i Frankrike,
PCP	Peptidylbærende protein

PCR	Polymerase kjedereaksjon
PEG	Polyetylenglykol
PKS	Polyketid Syntases
RDP	Recombination Detection Program
RNA	ribonukleinsyre
rpm	omdreining per minutt
rRNA	ribosomalt RNA
SDS	natriumdodecylsulfat
T-domenet	Tioleringsdomene
TE	Tioesterase
TEA	tris-eddiksyre-EDTA
tRNA	Transport RNA
UV	Ultrafiolett
X-GAL	5-brom-4-klor-3-indolyl- beta-D-galaktopyranosid

Adenosin = A  
Guanin = G  
Cytosin = C  
Tymidin = T

Alanin	Ala	A	GCG, GCA, GCT, GCC
Arginin	Arg	R	AGG, AGA, CGG, CGA, CGT, CGC
Asparagin	Asp	D	GAT, GAC
Aspartat	Asn	N	AAT, AAC
Cystein	Cys	C	TGT, TGC
Fenylalanin	Phe	F	TTT, TTC
Glutamin	Gln	Q	CAG, CAA
Glutamat	Glu	E	GAG, GAA
Glycin	Gly	G	GGG, GGA, GGT, GGC
Histidin	His	H	CAT, CAC
Isoleucin	Ile	I	ATA, ATT, ATC
Leucin	Leu	L	TTG, TTA, CTG, CTA, CTT, CTC
Lysin	Lys	K	AAG, AAA
Metionin	Met	M	ATG
Prolin	Pro	P	CCG, CCA, CCT, CCC
Serin	Ser	S	AGT, AGC, TCG, TCA, TCT, TCC
Treonin	Thr	T	ACG, ACA, ACT, ACC
Tryptofan	Trp	W	TGG
Tyrosin	Tyr	Y	TAT, TAC
Valin	Val	V	GTG, GTA, GTT, GTC

<b>FORORD .....</b>	<b>I</b>
<b>SAMMENDRAG .....</b>	<b>II</b>
<b>FORKORTELSER.....</b>	<b>III</b>
<b>1 INNLEDNING .....</b>	<b>1</b>
1.1 CYANOBAKTERIER – BLÅGRØNNALGER .....	1
1.2 IKKE-RIBOSOMAL PEPTIDSYNTETELSE .....	3
1.2.1 Adenyleringsdomenet (A-domenet) .....	6
1.2.2 Cyanopeptolin, Ahp, 3-amino-6-hydroksypiperidon .....	8
1.3 MÅL MED OPPGAVEN .....	11
<b>2 MATERIELL OG METODER.....</b>	<b>12</b>
DNA METODER.....	12
2.1.1 Bakteriekulturer.....	12
2.1.2 DNA isolering.....	12
2.1.3 Polymerase kjedereaksjon (PCR).....	13
2.1.4 Agarosegelelektroforese av DNA .....	15
2.1.5 Gelekstraksjon .....	16
2.1.6 Restriksjonskutting .....	17
2.1.7 Vektorsystem .....	18
2.1.8 Transformasjon.....	20
2.1.9 Kvikk-sjekk.....	22
2.1.10 Isolering av plasmid DNA.....	22
2.2 BIOINFORMATIKK .....	23
2.2.1 Sekvensanalyse .....	23
2.2.2 Fylogenetiske analyser.....	24
2.3 REKOMBINASJONSANALYSER.....	28
2.4 PROTEINMETODER .....	30
2.4.1 SDS-polyakrylamidgelelektroforese.....	30
2.5 HETEROLOG EKSPRESJON OG RENSING AV ENZYM.....	31
2.5.1 Induksjon med IPTG .....	31
2.5.2 Rensing av His-tag protein .....	31
2.6 ENZYMAKTIVITET .....	34
2.6.1 ATP-pyrofosfat utbytting ( $^{32}$ PPi/ATP utbytting). .....	34
<b>3 RESULTATER .....</b>	<b>36</b>
3.1 AHP DOMENET HOS CYANOBAKTERIER.....	36
3.1.1 PCR analyser.....	36
3.1.2 Sekvensanalyser.....	40

---

3.1.3	<i>Fylogenetiske analyser</i> .....	41
3.1.4	<i>Rekombinasjons analyser</i> .....	44
3.2	AHP DOMENEAKTIVITET .....	48
3.2.1	<i>Ekspresjon av Ahp-domenet fra stamme 172</i> .....	48
3.2.2	<i>PCR</i> .....	48
3.2.3	<i>Subkloning i pET21d-vektor</i> .....	49
3.2.4	<i>Transformasjon av E. coli BL21</i> .....	51
3.2.5	<i>Proteinekspresjon</i> .....	52
3.2.6	<i>Rensing av His-tag protein</i> .....	53
3.2.7	<i>ATP/PPi utbytting</i> .....	56
4	<b>DISKUSJON</b> .....	57
4.1	IDENTIFISERING AV AHP AKTIVERENDE DOMENER HOS CYANOBAKTERIER .....	57
4.2	IDENTIFISERING AV FORLØPEREN TIL AHP I CYANOPEPTOLIN .....	59
4.3	VIDERE FORSKNING .....	61
	<b>REFERANSER:</b> .....	63
5	<b>VEDLEGG</b> .....	70
5.1	SEKVENSAMMENTILLING AV A-DOMENER .....	70
5.2	SEKVENSERINGSPRIMERE .....	76
5.3	OPPSKRIFTER PÅ LØSNINGER.....	77

# Innledning

## 1.1 Cyanobakterier – blågrønnalger

Cyanobakterier er fotosyntetiske gramnegative bakterier, også kjent som blågrønnalger. Historisk sett ble de regnet som alger ettersom de utfører fotosyntese. Cyanobakterier er utbredt over hele jorda og omfatter ca 150 slekter med 1500 arter. De finnes i omtrent alle habitater med tilgang på lys, flesteparten i akvatiske habitater, men også i mer ekstreme habitater som ørken eller arktiske strøk. De er lite næringskrevende, bruker lys som energikilde, vann som elektrondonor og CO<sub>2</sub> som karbonkilde. Mange cyanobakterier har også evnen til å fikse nitrogen direkte fra atmosfæren.

Cyanobakterier er en stor variert gruppe som tradisjonelt har blitt klassifisert på grunnlag av morfologi, fysiologiske og økologiske forhold (tabell 1) (Castenholz, 2001). Forholdene som har vært grunnlaget for en slik klassifisering kan påvirkes mye av miljø og vekstforhold. Molekylær klassifisering av cyanobakterier er basert på 16S rRNA-gensekvenser fra renkulturer (Wilmotte og Herdmann, 2001), denne metoden gir grunnlag for ny klassifisering basert på arvemateriale, og kan skille mellom morfologisk like stammer.

**Tabell 1 Tradisjonell inndeling av cyanobakterier (Rippka *et al.*, 1979)**

<b><u>Gruppe I</u></b>	<b><u>Gruppe II</u></b>	<b><u>Gruppe III</u></b>	<b><u>Gruppe IV</u></b>	<b><u>Gruppe V</u></b>
<i>Microcystis</i> <i>Aphanocapsa</i> <i>Merismopedia</i>	<i>Chroococidiopsis</i> <i>Dermocarpella</i> <i>Pleurocapsa</i>	<i>Planktothrix</i> ( <i>Oscillatoria</i> ) <i>Pseudanabaena</i> <i>Tychonema</i> <i>Phormidium</i>	<i>Anabaena</i> <i>Nostoc</i> <i>Aphanizomenon</i> <i>Cylindrospermopsis</i>	<i>Chlorogloeopsis</i> <i>Fischerella</i>

Evolusjonært er cyanobakterier svært interessante fordi de var blant de tidligste kjente livsformer på jorda, og har vært involvert i dannelsen av jordens atmosfære. Dannelsen av oksygen ved cyanobakteriell fotosyntese bidrar fremdeles til å holde vår atmosfære i balanse. Cyanobakterier er kilde til et stort antall biologisk aktive produkter. Frem til i dag er omtrent 800 molekyler kjent, hvor flere er av farmasøytisk interesse (Schwarzer *et al.*, 2003).

I sommerhalvåret er elver og innsjøer ofte utsatt for såkalte algeoppblomstringer. Ofte er disse oppblomstringene store populasjoner av cyanobakterier (Carmichael *et al.*, 2001). Slike store oppblomstringer kan være et problem for drikkevannskilder fordi noen arter av cyanobakterier produserer toksiner som er farlige for dyr og mennesker. Det mest kjente toksinet er levertoksinet mikrocystin, det produseres ikke bare av *Microcystis*, men også av stammer fra andre slekter av cyanobakterier som *Anabaena*, *Planktothrix* og *Nostoc* (Sieber og Marahiel, 2003; Welker og Döhren, 2006). *Anabaena* er filamentøse cyanobakterier, de produserer heterocyster, som er spesialiserte nitrogenfikserende celler. Slekten *Microcystis* danner store kolonier av enkeltceller som er omgitt av slim. Cellene produserer ofte gassvakuoler, noe som gjør at koloniene kan flyte på overflaten av vann. Slekten *Planktothrix* (*Oscillatoria*) består av ugreinede trådlignende kolonidannende filamentøse cyanobakterier, med en glidende bevegelse. Ved bruk av molekylære teknikker har det blitt vist at flere stammer av cyanobakterier produserer toksiner som mikrocystin og andre peptider ved hjelp av ikke-ribosomale peptidsyntetaser (Ehrenreich *et al.*, 2005). I cyanobakterier, og spesielt hos *Microcystis* kan morfologisk like stammer ha ulike typer og varianter av peptider (Welker *et al.*, 2003).



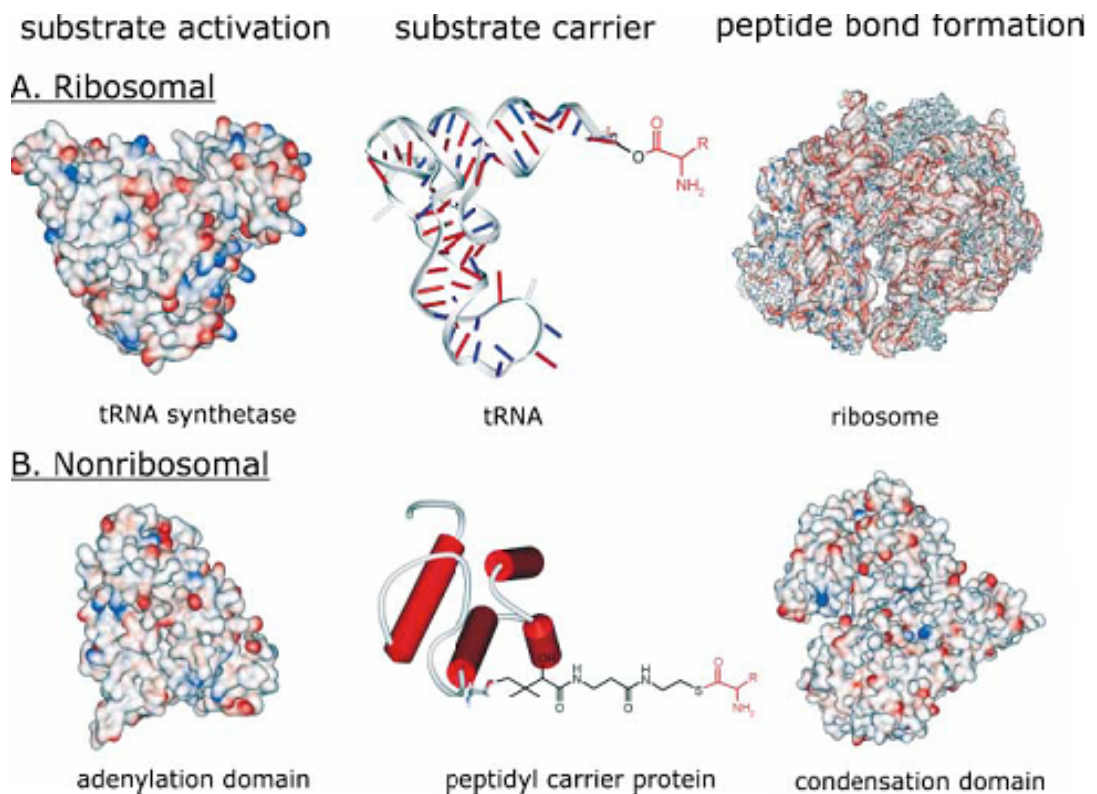
### 1.2 Ikke-ribosomal peptidsyntetese

Et stort antall biologisk aktive peptider syntetiseres ved ikke-ribosomal peptidsyntetase hos bakterier og sopp (Finking og Marahiel, 2004). Disse metabolittene har store funksjonelle og strukturelle forskjeller og inkluderer blant annet antibiotika, sideroforer (jernbærende molekyler), toksiner og immunosupprimerende midler.

Økt bruk av antibiotika har ført til at flere patogene bakteriestammer har blitt resistente mot antibiotika, dette har ført til økt interesse for små peptider i søken etter alternative medikamenter (Gold og Moellering, 1996).

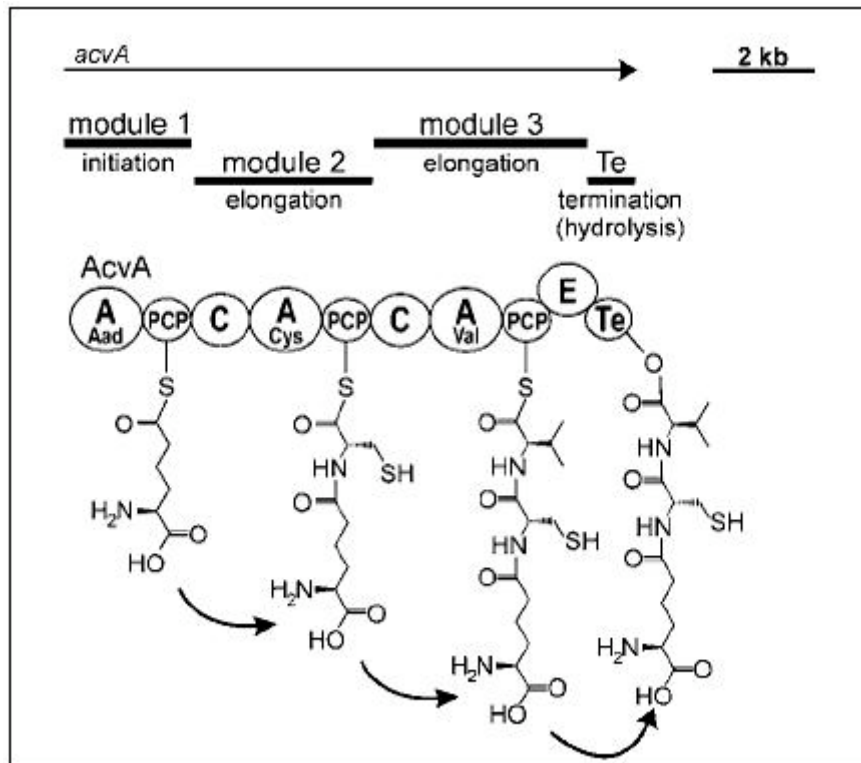
Ikke-ribosomale peptider dannes av store enzymkomplekser som er generert ribosomalt ved vanlig proteinsyntese. Aktivering av aminosyrer i en multienzymatisk prosess ligner mye på måten aminosyrene aktiveres på ved vanlig ribosomal peptidsyntese, men enzymene som er involvert er ikke strukturelt eller katalytisk like. Aktivering av aminosyrer i en ribosomal peptidsyntese er katalysert av tRNA syntetase, og produktet er aminoacyl-tRNA. I ikke-ribosomal peptidsyntese aktiveres substratet av adenyleringsdomenet (A-domenet) til hver modul, først ved dannelse av aminoacyladenylat, fulgt av bindingen av aminoacylgruppen som en tioester til tioleringsdomenet (peptidylbærende protein, PCP) i modulen. Dannelsen av peptidbindingene katalyseres av kondenseringsdomenet (C-domenet). I ribosomal syntese dannes peptidbindingene direkte av ribosomene, og mRNA fungerer som templat for å bestemme aminosyresekvensen til produktet.

Både det ribosomale og det ikke-ribosomale systemet har sine fordeler som gjør dem godt egnet til sine formål. Det ribosomale systemet har flere korrekturlesningsmekanismer som ikke er nødvendige i det ikke-ribosomale systemet. Det ikke-ribosomale systemet har klare fordeler i forhold til ribosomsystemet når det gjelder stor variasjon i strukturene til de ulike produktene. NRPS er ikke begrenset til 20 aminosyrer, men kan bruke flere hundre substrater som monomerer for peptidoppbygging (Sieber og Marahiel, 2003). Bakteriegener kodet av peptidsyntetaser, som er involvert i ikke-ribosomal syntese er samlet i operoner.



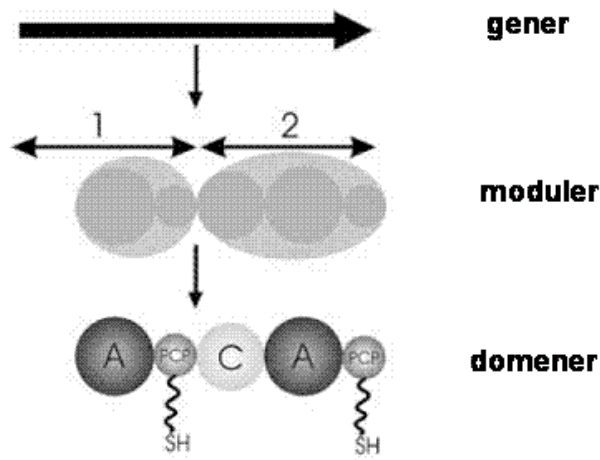
**Figur 1. Enzymer i det ribosomal og ikke-ribosomal peptid syntese systemet (Finking og Marahiel, 2004). A-domenet er strukturelt uten relasjon til t-RNA syntetase, men aktiverer også aminosyrer som acyladenylat. Aktivert aminosyrer blir transportert til kofaktoren til PCP, og C-domenet danner peptidbindinger mellom aminosyrene.**

Både ribosomal og ikke-ribosomal peptidsyntese består av aktivering av aminosyrer etterfulgt av dannelsen av peptidbindinger, dette repeteres til ønsket lengde på peptidkjeden er oppnådd. NRPS er organisert i grupper av aktive seter, kalt moduler. Hver modul består av flere domener. Hver modul aktiverer en aminosyre som blir innført i den voksende peptidkjeden. Rekkefølgen og antall moduler bestemmer vanligvis sekvensen og lengden på peptidet (Döhren *et al.*, 1997; Marahiel *et al.*, 1997).



**Figur 2. Organisering i NRPS (Mootz *et al.*, 2002).** De tre domenene er organisert i rekkefølgen, C ± A ± PCP i elongerings modulen som tilfører en aminosyre til den voksende peptidkjeden. Den første aminosyren til peptidkjeden blir inkorporert av initieringsmodulen som mangler C-domene. Under syntesen blir substratet kovalent bundet som tioester til fosfatpantetein til PCP.

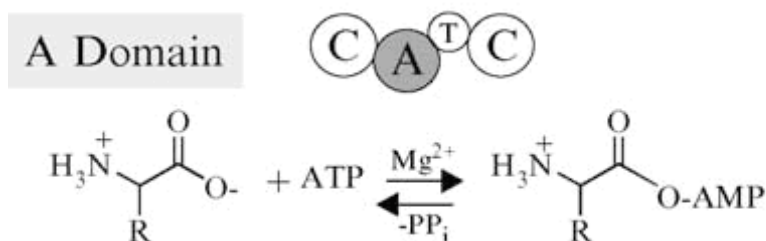
En modul vil alltid inneholde et adenyleringsdomene, et tioleringsdomene og et kondenseringsdomene. Adenyleringsdomenet (A-domenet) består av omtrent 550 aminosyrer og er ansvarlig for seleksjon og aktivering av aminosyresubstrat. Tioleringsdomenet (T-domenet) består av omtrent 80 aminosyrer, og er området hvor substratet blir kovalent knyttet til enzymet ved tioesterbinding til fosfopanteteingruppen. 4'-fosfopantetein er en viktig prostetisk gruppe i flere acylbærende proteiner, og fungerer som en svingende arm for å knytte substratet til enzymet. Kondenseringsdomenet (C-domenet) er omtrent 400 aminosyrer langt, og katalyserer dannelsen av peptidbinding mellom aminoacylgruppen og den frie aminogruppen på aminosyren som er bundet til neste T-domene. Modifiseringsaktiviteter som N-metylering og epimerisering er ganske vanlige i peptidsyntetaser. Tioesterase (TE)-aktiviteten katalyserer frigjøringen av produktet (Welker og Döhren, 2006).



**Figur 3. Gen, moduler, domener (Linne og Marahiel, 2004).** Gener er organisert i operoner, innen et gen er det moduler og hver modul er et definert område som er ansvarlig for inkorporering av en aminosyre i den voksende peptidkjeden. Moduler inndeles i domener, som er enzymatiske enheter som katalyserer skrittene for substrat aktivering, kovalent binding og peptidbånddannelse.

### 1.2.1 Adenyleringsdomenet (A-domenet)

For hver aminosyre i peptidproduktet finnes det et bestemt adenyleringsdomene, hvor rekkefølgen av disse domenene i operonet vanligvis bestemmer primærstrukturen til produktet (Marahiel *et al.*, 1997). De senere årene har det blitt gjort store fremskritt i oppdagelser og sekvensering av NRPS gener og operoner fra bakterier og sopp. Mange av disse peptidsyntetasene produserer peptider med ukjent struktur, men de har de samme konserverte signatursekvensene (Challis *et al.*, 2000).



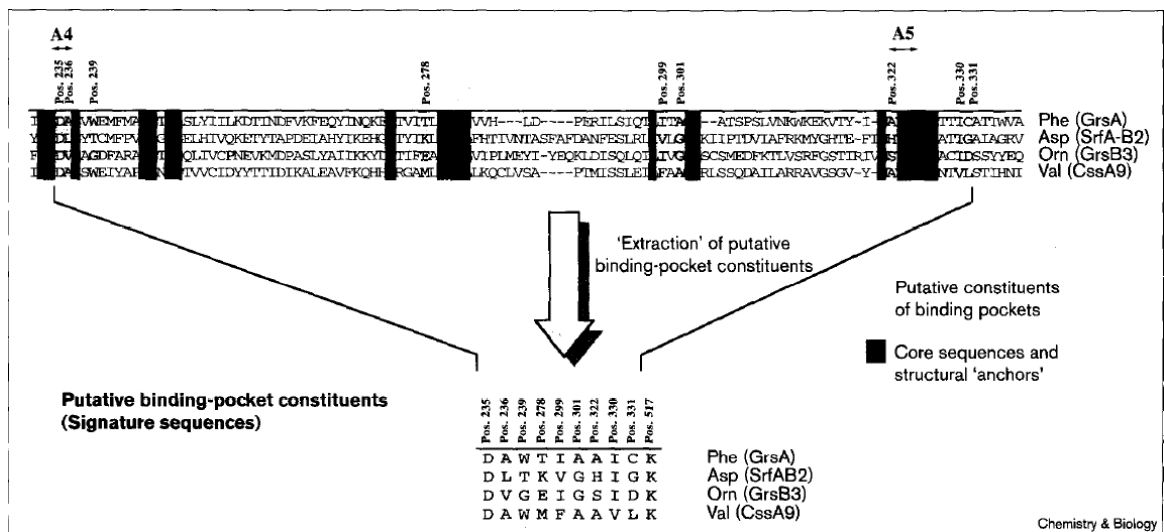
**Figur 4. Skjematisk fremstilling av reaksjonen som katalyseres av A-domenet (Linne og Marahiel, 2004).** A-domenet er ansvarlig for seleksjon og aktivering av aminosyresubstrat.

Fra cyanobakterier har omtrent 200 A-domener til nå blitt sekvensert. Disse er integrert i NRPS-systemer. Ved sekvenssammenligning kan man finne 10 konserverte områder (A1-A10) (Tabell 2) (Marahiel *et al.*, 1997). Fylogenetiske analyser viser at det er størst likhet mellom A-domener som aktiverer samme substrat.

**Tabell 2 Konserverte områder i adenyleringsdomenet (Marahiel *et al.*, 1997)**

Navn	Konsensussekvens
A1	L(TS)Y <sub>x</sub> EL
A2	LKAG <sub>x</sub> AYL(VL)P(LI)D
A3	LAY <sub>xx</sub> YTSG(ST)TG <sub>x</sub> PKG
A4	FD <sub>x</sub> S
A5	N <sub>x</sub> YGPTE
A6	GEL <sub>x</sub> IXG <sub>x</sub> G(VL)ARGYL
A7	Y(RK)TGDL
A8	GR <sub>x</sub> D <sub>x</sub> QVKIRG <sub>x</sub> RIELGEIE
A9	LP <sub>x</sub> YM(IV)P
A10	NGK(VL)DR

Strukturbaserte undersøkelser har identifisert 10 aminosyreposisjoner som bidrar til å bestemme bindingssetets spesifisitet for aminosyrer. Disse signatursekvensene blir kalt NRPS koden. Figur 5 viser hvordan sekvenser av A-domener kan sammenstilles, for så utfra gitte posisjoner bestemme signatursekvensen. Posisjonene er bestemt utfra krystallstrukturen til det fenylalaninaktiverende domenet i gramicidin S (GrsA) syntetase (Stachelhaus og Marahiel, 1995), og brukes som generell strukturmodell for å forutsi A-domene-substrataktivering (Stachelhaus *et al.*, 1999).



**Figur 5. Identifisering av signatursekvenser (Stachelhaus *et al.*, 1999).**  
 Ved sekvenssammenstilling identifiserer 10 aminosyreposisjoner som antyder hvilken aminosyre A-domenet aktiverer.

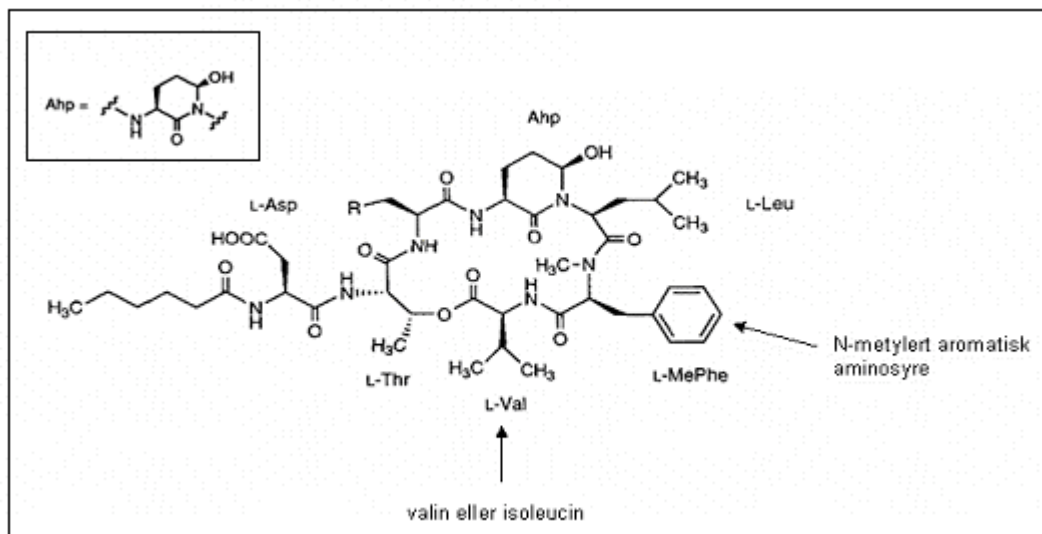
De fleste ikke-ribosomale peptider fra prokaryoter og lavere eukaryoter er klassifisert som sekundære metabolitter, dvs. at de sjelden har en rolle i primærmetabolismen, vekst eller reproduksjon, men gir muligens en fordel for den produserende organismen. Biosyntese av ikke-ribosomale peptider krever store deler av cellens energi og næring, men de biologiske funksjonene til produktene er enda ikke kjent.

### 1.2.2 Cyanopeptolin, Ahp, 3-amino-6-hydroksypiperidon

Cyanobakterier syntetiserer en rekke biologisk aktive peptider. Sykliske strukturer, og uvanlige og modifiserte aminosyrer, tilsier at mange av disse peptidene er syntetisert ved ikke-ribosomal peptidsyntese ved hjelp av tiotemplatmekanismen (Marahiel *et al.*, 1997). Generelt er sykliske peptider og depsipeptider de vanligste strukturene, som f. eks mikrocystin og cyanopeptolin. Et depsipeptid er en peptidforbindelse med både ester- og peptidbindinger. Mikrocystin er det best studerte peptidet. Det produseres ikke bare av *Microcystis*-arter men også av arter fra *Anabaena*, *Nostoc* og *Planktothrix* (*Oscillatoria*). Cyanopeptolin er etter mikrocystin det vanligste peptidet som produseres av *Microcystis*-arter. Selv om det er mange sykliske peptider hos cyanobakterier, har ganske få klasser av NRPS operoner til nå blitt karakterisert. De mest omfattende undersøkelsene har blitt gjort på mikrocystin

syntetase-genklustere (mcy, nda), hvor klusteret i to stammer av *Microcystis* (Nishizawa *et al.*, 2000), en *Planktothrix* (Christiansen *et al.*, 2003), en *Anabaena* (Rouhiainen *et al.*, 2004) og en *Nodularia* (Moffitt og Neilan, 2004) har blitt karakterisert.

Cyanopeptoliner utgjør en klasse av proteaseinhiberende sykliske peptider som inneholder aminosyrederivatet 3-amino-6-hydroksypiperidon, Ahp.



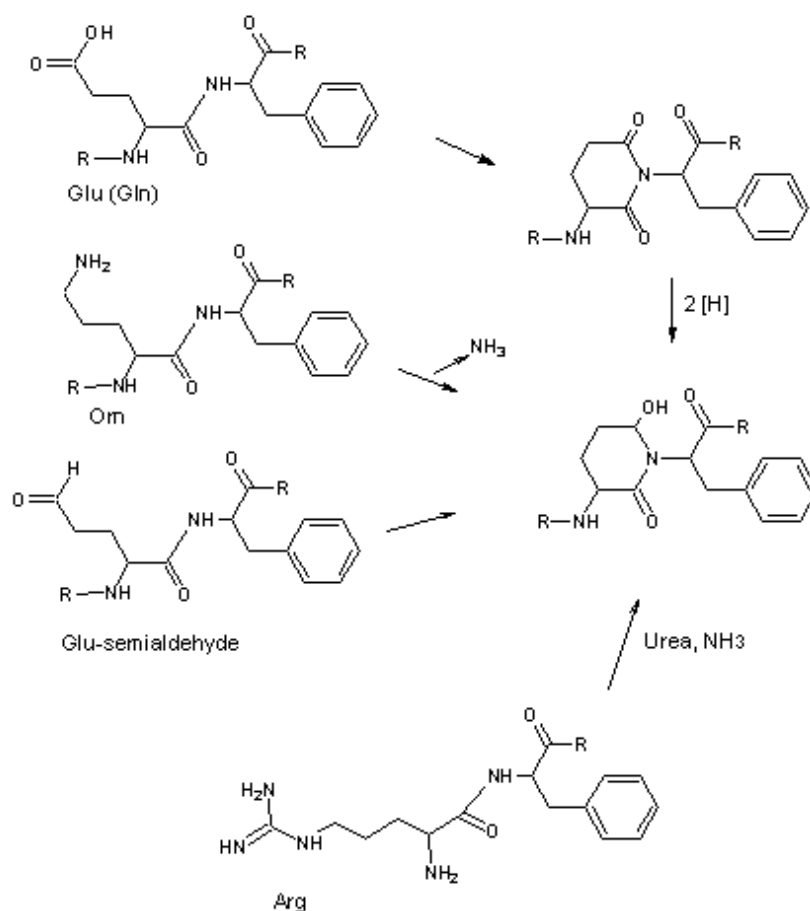
**Figur 6. Skjematisk struktur av cyanopeptolin isolert fra *Microcystis aeruginosa* PCC 7806 (Namikoshi og Rinehart, 1996). Cyanopeptolin består av syv aminosyrer, hvor seks danner en ring. Alle cyanopeptoliner inneholder 3-amino-6-hydroksy-2-piperidone, Ahp.**

I litteraturen har depsipeptider med den samme basale strukturen fått flere navn. Hos *Microcystis* er de blitt kalt cyanopeptolin (Martin *et al.*, 1993), mikropeptin (Okino *et al.*, 1993), aeruginopeptin (Harada *et al.*, 1993) eller mikrocystilid (Tsukamoto *et al.*, 1993). Hos *Anabaena* har anabaenopeptilid samme struktur (Rouhiainen *et al.*, 2000). Cyanopeptolin fra *Microcystis* var den første varianten som ble beskrevet fullstendig, og dette har de siste årene blitt den mest brukte fellesbetegnelsen på denne klassen av peptider.

Alle kjente varianter av depsipeptider som inneholder Ahp hos *Microcystis* har enten isoleucin eller valin som C-terminal aminosyre, alle aminosyrene er i L-konfigurasjon, sidekjeden til den N-terminale aminosyren er koblet til en kortkjedet karboksylsyre med eller uten andre funksjonelle grupper, og aminosyren ved siden

den N-terminale aminosyren er en N-metyltert aromatisk aminosyre (Bister *et al.*, 2004).

Forløperaminosyren aktiveres som en aminoacyladenylat av Ahp A domenet til peptidsyntetasen, og blir overført som en tioester til en fosfopantetein gruppe på et nabo peptidylbærende (P) domene. Ahp innføres i den voksende peptidkjeden fra en ukjent forløper, mest sannsynlig glutamat, glutamin, glutaminsyresemiaraldehyd, arginin eller ornitin (Figur 7). For anabaenopeptilid har det blitt foreslått at det er glutamin som er den ukjente forløperen (Rouhiainen *et al.*, 2000). Cyanopeptolinsyntetaser er til nå beskrevet på gennivå for *Anabaena* 90 (adp) (Rouhiainen *et al.*, 2000), *Microcystis* NC 172/5 (mcn) (Tooming-Klunderud *et al.*, 2007) og *Planktothrix* NIES 205 (oci) (Rounge *et al.*, 2007).



**Figur 7. Mulige Ahp forløpere (Tom Kristensen).  
Glutamat/glutamin, ornitin, glutaminsyresemiaraldehyd og arginin.**



## 1.3 Mål med oppgaven

Formålet med denne masteroppgaven var å lære mer om A-domenet som aktiverer Ahp i cyanopeptolin.

For å gjøre dette var første del av oppgave å finne Ahp-aktiverende domener i flere stammer av cyanobakterier, ved bruk av konsensus-PCR-primere som skiller mellom Ahp A-domener og A-domener som aktiverer andre aminosyrer. Sekvensene fra de ulike stammene vil gi informasjon om evolusjon og fordeling av Ahp-domener. Stammer med Ahp-domener lager trolig cyanopeptolin.

Reaksjonen som fører til dannelse av Ahp, den uvanlige aminosyren i cyanopeptolin, er ikke kjent. Andre del av oppgaven var å identifisere forløperen for Ahp-delen i cyanopeptolin hos NC 172 *Microsystis cf. wesenbergii*, ved å finne ut hvilken aminosyre Ahp A-domenet aktiverer. Dette har tidligere blitt gjort for andre A-domener ved å uttrykke domenet i *E. coli* som his-tag-proteiner, for så å studere aktiveringsprosessen ved  $^{32}\text{PPi}/\text{ATP}$  utbytting.

## 2 Materiell og Metoder

Hver metode som er brukt i denne oppgaven, og noe teori beskrives i dette kapitlet, oppskrifter på løsninger finnes i vedlegg 5.3. og primere brukt til sekvensering finnes i vedlegg 5.2

### DNA metoder

#### 2.1.1 Bakteriekulturer

Bakteriestammer brukt i denne oppgaven kommer fra kulturer til Norsk Institutt fra Vannforskning (NIVA), (Tabell 5). Det er i tillegg brukt to stammer fra Pasteurinstituttet i Frankrike, PCC 7806 og PCC 7319.

#### 2.1.2 DNA isolering

DNA var isolert tidligere fra en del av kulturene av Bjørg Mikalsen og Ave Tooming-Klunderud ved bruk av en metode for isolering av plante-DNA (Mikalsen *et al.*, 2003).

Fra de øvrige kulturene ble genomisk DNA isolert fra kulturene ved hjelp av Wizard Genomic DNA purification kit fra Promega ([www.promega.com](http://www.promega.com)). Wizard Genomic DNA purification kit er laget for å isolere DNA fra gram negative bakterier. Cellene ble lysert, og uløselige proteiner og cellerester fjernet ved sentrifugering. Videre ble RNA fjernet med enzymet RNase, og løselige proteiner fjernet ved å tilsette protease. Tilslutt ble genomisk DNA oppkonsentrert og avsaltet ved isopropanol-presipitering. Konsentrasjoner ble målt med NanoDrop<sup>®</sup> ND-1000 UV-Vis Spektrofotometer ([www.nanodrop.com](http://www.nanodrop.com)), og prøvene ble lagret ved -20 °C.

En annen metode brukt for DNA isolering er å lysere cellene alkalisk, for videre å bruke frigjort DNA direkte til PCR. Polyetylenglycol (PEG) med høy konsentrasjon

(>50 %) og alkalisk pH blir brukt til å lysere cellene (Chomczynski og Rymaszewski, 2006). Nærvær av PEG gjør det mulig å oppnå en pH høyere enn 13 med bare 20mM KOH. Det er nødvendig med pH over 13 for å holde DNA i løsningen, ved lavere pH kan DNA presipitere. Denne lave KOH-konsentrasjonen gjør det mulig å tilsette et alkalisk bakterielysat direkte til en PCR-reaksjonsblanding uten at pH i den endelige PCR-blandingen endres vesentlig. Alkalisk PEG brukt i denne oppgaven består av 60 % PEG, KOH og vann og har en pH på 13,5.

Prøvene fra NIVA blir levert i kulturer, de blir spunnet ned med isopropanol og cellene blir resuspendert med isopropanol og lagret ved -20 °C. Når DNA skal isoleres ble prøvene sentrifugert, isopropanol fjernet og PEG tilsatt. Dette blir inkubert i ca 15min ved romtemperatur, blandet godt, og 3-5 µl ble brukt direkte til PCR.

### 2.1.3 Polymerase kjedereaksjon (PCR)

PCR er en metode for å amplifisere en bestemt DNA-sekvens *in vitro*. PCR reaksjonen foregår ved at en blanding bestående av DNA-templat, et primerpar, dNTP, og en termostabil polymerase eksponeres for ulike temperaturer i flere sykluser. Reaksjonen starter ved at den doble DNA-tråden denatureres og deles til to enkelttråder ved høy temperatur. Temperaturen senkes til en temperatur som er optimal for at primerene skal binde seg til templat-DNA, det må være to primere som hybridiserer til hver sin komplementære DNA-tråd. Primerene definerer området av DNAet som skal kopieres, og fungerer som startsted for DNA polymerasen. Så økes temperaturen til det som er optimalt for at DNA polymerasen skal syntetisere nytt DNA ved 68-72 °C. Ved at de ulike trinnene i PCR gjentas i flere omganger skjer kopiering av DNA eksponensielt.

I denne oppgaven er det brukt Advantage 2 polymerase fra Clontech ([www.clontech.com](http://www.clontech.com)) og Phusion polymerase fra Finnzymes ([www.finnzymes.fi](http://www.finnzymes.fi)). Primere som er brukt er vist i tabell 3. PCR-programmet ble optimalisert ved å kjøre temperaturgradient på 55 til 65 °C med ulike primerpar, og programmet som ga best resultat og ble brukt videre (Figur 8).

95 °C 1min	} x 30
95 °C 15 sek	
55 °C 1min	
68 °C 2min	
68 °C 5min	
4 °C ∞	

**Figur 8. PCR program optimalisert for denne oppgaven**

For å gjøre det enklere å subklone PCR produktet inn i en vektor tilfører polymerasen en adenin (A) på 3'ende til produktet. Vektoren som brukes har en 3' T (tymin) overheng, som gjør at det er enklere å ligere PCR produktet inn i vektoren. Advantage polymerasen og Phusion polymerase har korrekturlesningsaktivitet som reduserer forekomsten av polymeraseinnførte feil i produktet, men som også i en viss grad vil fjerne A-overheng. Det vil allikevel være tilstrekkelig med 3'-A overheng til at ligering til TA-vektor skjer effektivt.

### Degenererte konsensusprimere

Primere som er brukt i denne oppgaven (tabell 3) er utformet av professor Tom Kristensen. Ahp-primere er utformet ved å sammenligne konserverte områder i kjente Ahp-domener fra stammer av *Microcystis*, *Planktothrix* og *Anabaena*.

Tabell 3. Primere brukt til PCR

Primer	Sekvens	
Ahp1F	5'- GA(CT) CC(AGCT) GA(CT) TA(CT) (AG)C(AGCT) CA(AG) CA -3'	Ahp primer
Ahp2F	5'- TT(CT) GC(AGCT) TC(AGCT) TT(CT) GG(AGCT) TT(CT) GA(CT) GT -3'	Ahp primer
Ahp7R	5'- TG(AG) TT(AG) TC(AG) TT(AGCT) CC(AG) AA(AG) CA -3'	Ahp primer
T3-Ahp1F	5'- ATT AAC CCT CAC TAA AGG GAG A(CT)C C(AGCT)G A(CT)T A(CT)(AG) C(AGCT)C A(AG)C A -3'	Ahp primer
T3-Ahp2F	5'- ATT AAC CCT CAC TAA AGG GAT T(CT)G C(AGCT)T C(AGCT)T T(CT)G G(AGCT)T T(CT)G A(CT)G T -3'	Ahp primer
T7-Ahp7R	5'- TAA TAC GAC TC ACTA TAGG GT G(AG)T T(AG)T C(AG)T T(AGCT)C C(AG)A A(AG)C A -3'	Ahp primer
Ahp_172F	5'- CCT ACC CAT GGC ATT GGT TGT TA ACCT GCA -3'	ekspressjon av Ahp
Ahp_172R	5' – TTG TAA CTC GAC TTT TTG CCA AAC TGC AAA -3'	ekspressjon av Ahp
mcnE-F	5'- GGA GGC AAT TGG TCC ATG GAA GTA GGA GAT -3'	ekspressjon av mcnE
mcnE-R	5' – AAC GGG TTT GGT ATC AAA TCT CGA GGC AAA – 3'	ekspressjon av mcnE

### 2.1.4 Agarosegelelektroforese av DNA

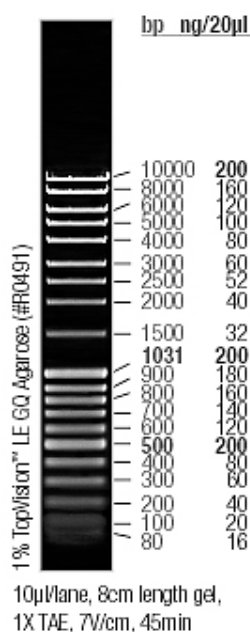
Elektroforese av DNA ble utført i en agarosegel. Dette er en metode for å separere DNA-fragmenter med ulike størrelser.

Agarosegeler ble laget ved å blande agarosepulver med bufferløsning, løse det opp ved oppvarming, og la løsningen stivne i en passende form. Når gelen stivner dannes det et nettverk der tettheten er bestemt av konsentrasjonen til agarosen. Ved å variere konsentrasjonen av agaroseløsningen fra 0,5- 2 % får man geler med ulike porestørrelser som har ulike separasjonsområder. DNA-prøvene ble applikert i brønner på gelen, og deretter plassert i et elektrisk felt. DNA er negativt ladet og vil vandre i gelen mot den positive polen (anoden). Små fragmenter vandrer raskere enn store, og for å bestemme størrelsen på fragmentene kjører man samtidig flere markører med kjent størrelse.

For å visualisere DNA-fragmentene tilsetter man etidiumbromid (EtBr). Etidiumbromid (EtBr) er en fluoriserende interkalerende stoff som brukes til merking

av nukleinsyrer i forbindelse med gelelektroforese. Molekylet har tilnærmet lik størrelse som et basepar i DNA og kan legge seg mellom baseparene i heliksen. Etter behandling blir nukleinsyrer synlige som bånd på gelen ved belysning med UV-lys.

Her er det for å lage en gel brukt 50 ml 1 % agarose løst i 1x TEA-buffer (Tris-acetat-EDTA) og tilsatt 2,5 µl 1 % etidiumbromid (Quantum). DNA prøvene er tilsatt fargebuffer (6x loading dye, Fermentas), og applisert på gelen sammen med markører (MassRuler™, Fermentas, se figur 9). Gelen ble kjørt i ca 1 time med konstant spenning på 100 volt før den ble avbildet.



**Figur 9. MassRuler™ DNA markør, Fermentas.**

### 2.1.5 Gelekstraksjon

DNA fragmentene ble rensset med E.Z.N.A® Gel Extraction Kit fra Omega Bio-Tek ([www.omegabiotek.com](http://www.omegabiotek.com)). Kittet baserer seg på spinnkolonneteknologi der DNA-fragmenter kan separeres fra en løsning ved hjelp av bindingsegenskapene fragmentene har til en silikagel-membran. DNA vil binde til membranen og kontaminanter kan fjernes via flere vaskeprosesser med ulike buffere. Siden binding av DNA til membranen avhenger av nærvær av kaotrope salter vil DNA elueres ved tilføring av en elueringsbuffer med lav ionestyrke.

DNA fragmenter ble kuttet ut av agarosegeler, veid og tilsatt bindingsbuffer fra kit'et. Gelbitene ble smeltet ved 55-60 °C, og løsningen ble overført til en kolonne som består av DNA-bindende silikamembran. DNA ble bundet til membranen og kontaminanter ble fjernet ved flere vaskeprosesser med ulike buffere fra kit'et. DNA ble tilslutt eluert med nukleasefritt vann.

DNA konsentrasjoner ble målt med NanoDrop® ND-1000 UV-Vis Spektrofotometer fra NanoDrop Technologies ([www.nanodrop.com](http://www.nanodrop.com)).

### 2.1.6 Restriksjonskutting

Restriksjonsendonukleaser er enzymer som kutter dobbeltrådet DNA. Enzymene binder spesifikt til restriksjonssetet som er den spesielle sekvensen av DNA, 4-8bp, ofte palindromer, som restriksjonsenzymet gjenkjenner og kutter. Noen restriksjonsenzymer kutter begge trådene nøyaktig i symmetriaksen og danner DNA-fragmenter med butte ender, såkalte "blunt-ends". Andre kutter hver tråd på likt sted og motsatt side av symmetriaksen slik at det dannes overhengende ender, såkalte klebrige ender, "sticky-ends". Ulike restriksjonsenzymer stiller forskjellige krav til optimale kuttebetingelser. Hensyn til temperatur og buffertype er av avgjørende betydning for å oppnå vellykket kutting. Buffertype ble valgt etter enzymleverandørens anbefalinger.

**Tabell 4 Restriksjonsenzymer brukt i denne oppgaven:**

Enzym	Kuttesekvens	Buffer	Leverandør
NcoI	CCATGG	Buffer Tango	Fermentas
AvaI (Eco88I)	CTCGAG	Buffer Tango	Fermentas
PstI	CTGCAG	Buffer O	Fermentas
ClaI	ATCGAT	NEBuffer4 + BSA	NEB
HindIII	AAGCTT	NEBuffer2	NEB

Reaksjoner ble satt opp med 200-1000 ng DNA som skulle kuttes, 2 µl buffer (tabell 4), 0,5-1µl enzym (Tabell 4) og vann til et totalvolum på 20 µl. Reaksjonsløsningene ble inkubert i 1 time ved 37 °C, og inaktivert ved 68 °C i 20 min.

## 2.1.7 Vektorsystem

DNA fragmenter kan liggeres inn i en egnet vektor ved hjelp av DNA ligase. Her ble det brukt to ulike ligeringsblandinger avhengig av hvilken vektor som ble brukt. For DNA som skulle klones ble det brukt pGEM-T Easy vektorsystem. For DNA som ble klonet for så å uttrykkes heterologt ble det benyttet Ligate-IT Rapid Ligation Kit med pET-vektor.

### 2.1.7.1 pGEM-T Easy vektor

pGEM-T Easy vector system fra Promega ([www.promega.com](http://www.promega.com)) er egnet for TA-kloning av PCR-produkter (Figur 10). Vektoren er lineær og har en 3'-T overheng i begge ender, slik at ligering av PCR produkt med 3'-A overheng kan oppnås. PCR-produktet tilsettes ligeringsbuffer, vektor og ligase, og inkuberes over natten før kompetente bakterier transformeres med ligeringsblandingen.

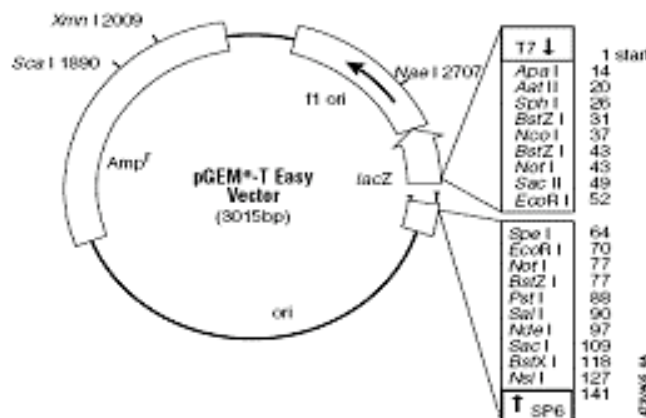
### Ligeringsblanding med pGEM-T Easy vektor system

5 µl 5x rapid ligation buffer  
1 µl (50ng) pGEM-T Easy vektor  
x µl PCR produkt  
2 µl Kontroll DNA  
1 µl T4 DNA ligase  
vann opp til totalt volum 10µl

Mengde PCR produkt beregnes utfra følgende forhold:

$$\frac{\text{ng vektor} \times \text{kb str innsatt}}{\text{kb str vektor}}$$

Ligeringsreaksjonen ble inkubert 1 time ved romtemperatur, eller ved 4 °C over natten.

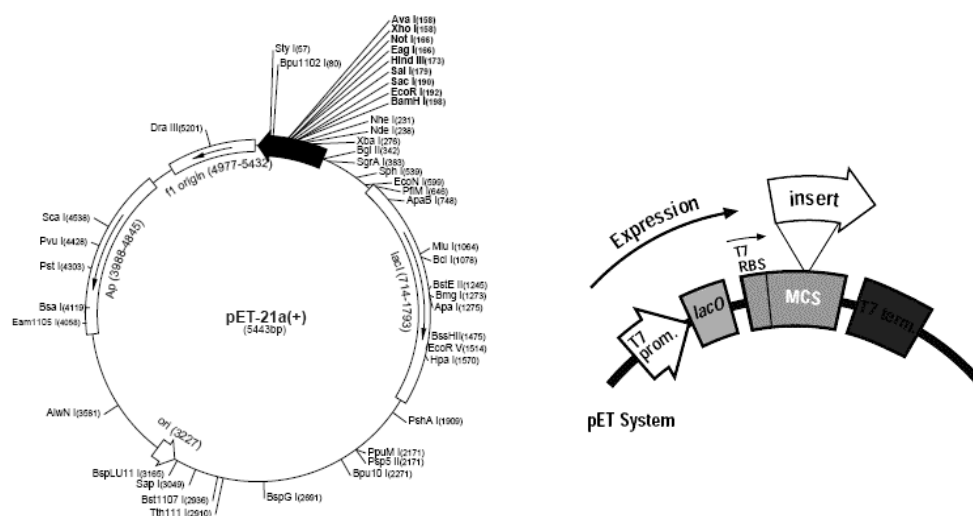


Figur 10. Kloningsvektor pGEM-T Easy Vector (3015bp) fra Promega.



### 2.1.7.2 pET-vektor

pET-vektorer er egnet til kloning og ekspresjon for rekombinante proteiner i *E. coli*. Her ble det brukt pET21d vektor som er basert på et T7-promotor drevet system (Figur 11). Det ønskede genet klones inn i pET-vektoren under kontroll av en T7 promoter, og ekspresjon skjer ved tilførsel av T7 RNA polymerase fra vertscellen. pET21d har en 6xhis-tag i C-enden av proteinet, noe som gjør det enklere å rense proteinet etter ekspresjon.



**Figur 11. pET21-vektor fra Novagen (5440bp), brukt til kloning og ekspresjon. pET21d har ampicillin resistens, og i kloningsste (MCS) har den blant annet kuttesete for NcoI og Aval som ble brukt i denne oppgaven.**

PCR-produktet ble kuttet med restriksjonsenzymmer og ligert til vektoren ved bruk av Ligate-IT Rapid Ligation Kit fra USB ([www.usbweb.com](http://www.usbweb.com)).

#### Ligeringsblanding med Ligate-IT Rapid Ligation Kit

- 500 ng plasmid
- 50-100 ng vektor
- 4 µl 5x ligate-it buffer
- 1 µl ligate-it T4 DNA ligase
- totalt 20 µl

Ligeringsreaksjonen ble inkubert i 7 min ved romtemperatur.

### 2.1.8 Transformasjon

Transformasjon er en prosess for å få et plasmid/ en ekspresjonsvektor inn i celler, slik at det ønskede plasmidet kan bli replikert og det innsatte gen eventuelt uttrykt. Det er vanlig å bruke *Escherichia coli*-celler til dette, ettersom de vokser raskt og er nøyne studert tidligere. Hvilken *E. coli*-stamme man velger avhenger av formålet. For oppformering av plasmid ble *E. coli* DH5 $\alpha$  brukt, og for transformasjon av plasmid for proteinekspresjon ble *E. coli* BL21(DE3)LysS brukt. For at cellene skal kunne ta opp DNA må de gjøres kompetente. Det finnes ulike metoder på å gjøre cellene kompetente. Her ble Inoue-metoden brukt for tillaging av kompetente *E. coli* DH5 $\alpha$  celler (Inoue, 1990), og CaCl<sub>2</sub> metoden for *E. coli* BL21 celler (Sambrook og Russell, 2001).

#### 2.1.8.1 Inoue-metoden for tillaging av kompetente *E.coli* DH5 $\alpha$

*E. coli* DH5 $\alpha$  ble strøket ut på en LB-skål, og inkubert over natt ved 37 °C. Neste dag ble noen kolonier fra skålen overført til 100 ml SOB-medium, og inkubert ved 37 °C i 3 timer. OD ble målt, og en beregnet mengde av kulturen til OD<sub>600</sub>=0,05 ble overført til 250 ml SOB, og inkubert ved 18 °C overnatt. Celletettheten ble målt neste dag, og kulturen avkjølt på is i 10min før den ble sentrifugert ved 2500 x g i 10min ved 4 °C. Supernatanten ble fjernet og cellene resuspendert i 80 ml iskald TB-buffer. Etter 10min på is ble cellene sentrifugert på nytt som beskrevet, og cellene resuspendert i 20 ml TB. Så ble 3,5 % DMSO tilsatt. Dette ble avkjølt på is i 5 min før det ble tilsatt 7 % DMSO, og satt på is i 10min. Cellene ble sjokkfrosset i flytende nitrogen, og oppbevart ved -70 °C til senere bruk.

#### ▪ Transformasjon med *E. coli* DH5 $\alpha$

Inoue-metoden for tillaging av kompetente celler ble brukt og cellene ble lagret ved -70°C. Til transformasjon ble 50  $\mu$ l av cellene overført til 3  $\mu$ l av ligeringsreaksjonen, mikset forsiktig og satt på is i 20 min. Deretter ble cellene varmesjokket på 42 °C i 45 sekunder, og ble avkjølt igjen på is i 2 min. De ble så strøket ut på LB skåler tilsatt ampicillin (100  $\mu$ g/ml), IPTG (isopropyl-beta-D-tiogalaktopyranoside, 1 mM) og X-Gal (5-brom-4-klor-3-indolyl- beta-D-galaktopyranosid, 20  $\mu$ g/ml). Skålene ble inkubert over natt ved 37 °C.

For å se om kloningen har vært vellykket har pGEM-T Easy vektorsystemet i tillegg til antibiotikaseleksjon, blå/hvit-seleksjon ved transformasjon. PCR-produktet blir satt inn i vektoren i lacZ genet. Dette genet koder for  $\beta$ -galaktosidase som kutter X-gal til et farget produkt som gir blå kolonier. IPTG induserer syntese av  $\beta$ -galaktosidase. Celler som tar opp innskuddet forstyrrer den kodene sekvensen til  $\beta$ -galaktosidase, slik at positive kolonier blir hvite (Sambrook og Russell, 2001).

### 2.1.8.2 CaCl<sub>2</sub>-metoden for kompetente *E. coli* BL21

*E. coli* BL21(DE3)LysS ble strøket ut på en LB-skål med kloramfenikol, og inkubert over natt ved 37 °C. Neste dag ble kolonier fra skålen inokulert med 100 ml LB med kloramfenikol til OD<sub>600</sub> var omtrent 0,3. Deretter ble cellene avkjølt på is i 10 min, for så å bli sentrifugert på 5000 x g i 10 min ved 4 °C. Supernatanten ble fjernet og cellene resuspendert i 10 ml iskald 100mM CaCl<sub>2</sub>. Etter 5min på is ble cellene sentrifugert på nytt som beskrevet over, supernatanten fjernet og cellene resuspendert i 2 ml 100 mM CaCl<sub>2</sub>. Cellene stod stå på is i minst en time, før de ble brukt til transformasjon.

#### ▪ Transformasjon med *E. coli* BL21

*E. coli* BL21-celler har det ikke vært mulig å gjøre kompetente ved Inoue-metoden, så disse må gjøres kompetente rett med CaCl<sub>2</sub>-behandling før bruk. *E. coli* BL21 celler er godt egnet for proteinekspresjon, fordi de inneholder et T7 RNA polymerase gen som er under kontroll av en lac-promotor. Her er det brukt *E. coli* BL21(DE3)LysS, LysS plasmidet koder for T7 lysozym, som er en naturlig inhibitor av T7 RNA polymerase, og derfor hindrer genet i å bli transkribert i fravær av indukere som IPTG. Ved tilsetning av IPTG til en voksende *E. coli* BL21 cellekultur vil T7 RNA polymerasegenet induseres.

Til transformasjon ble 200  $\mu$ l kompetente celler overført til 3  $\mu$ l av ligerinsreaksjonen. Dette ble mikset forsiktig og satt på is i 20 min. Cellene ble varmesjokkbehandlet ved 42 °C i 45 sekunder, og avkjølt på is i 2 min. 100  $\mu$ l ble strøket ut på LB skåler tilsatt ampicillin (100  $\mu$ g/ml) og kloramfenikol (25 mg/ml). Skålene ble inkubert over natt ved 37 °C.

### 2.1.9 Kvikk-sjekk

En effektiv metode for å sjekke om transformasjonen har vært vellykket er å lysere cellene og ekstrahere DNA som kan visualiseres på en agarosegel. I denne oppgaven ble det brukt 100 µl av overnattkultur tilsatt 8 µl loading buffer og 50 µl fenol:kloroform (1:1). Prøvene ble mikset i 30 sek og sentrifugert i 3 min. 30 µl av den øverste fasen, vannfasen, ble applisert på en agarosegel. Proteinene i cellene felles ut ved hjelp av fenol:kloroform, mens nukleinsyrene blir værende i vannfasen. I tillegg til en molekylvektsmarkør brukes også ligert vektor uten innskudd som kontroll på gelen. Plasmider med innskudd vil danne et bånd på gelen som er høyere enn vektor uten innskudd, det vil også være sterke bånd av kromosomalt og ribosomalt RNA (Akada, 1994).

### 2.1.10 Isolering av plasmid DNA

Isolering av plasmid-DNA ble utført med Wizard mini/midi preps DNA Purification System fra Promega ([www.promega.com](http://www.promega.com)). Et utvalg av positive enkeltkolonier fra skålene dyrkes over natten i LB-media med passende antibiotika seleksjonsmiddel på 37 °C og det produseres et lysat som overføres til en kolonne med en silikamembran. Kittet baserer seg på spinn-kolonneteknologi der DNA-fragmenter kan separeres fra en løsning ved hjelp av bindingsegenskapene fragmentene har til silikagel-membran. DNA vil binde til membranen og kontaminanter kan fjernes ved vaskeprosesser med ulike buffere. Siden DNA bare bindes til membranen i nærvær av kaotrope ioner, vil DNA elueres ved tilføring av en elueringsbuffer med lav ionestyrke eller vann.

Konsentrasjonen av DNA ble målt med NanoDrop<sup>®</sup> ND-1000 UV-Vis Spektrofotometer ([www.nanodrop.com](http://www.nanodrop.com)). For videre lagring ble bakteriekulturene tilsatt 60 % glycerol og frosset ned med flytende nitrogen, og lagret ved -80 °C (Sambrook og Russell, 2001).

## 2.2 Bioinformatikk

### 2.2.1 Sekvensanalyse

DNA sekvensering er en teknikk der man får bestemt baserekkefølgen i et DNA-molekyl. Det finnes flere metoder for sekvensering av DNA, men Sangers kjedeterminerings-metode er den mest benyttede sekvenseringsmetoden i dag (Sanger *et al.*, 1977). Sekvensering ble utført på ABI laboratoriet til CEES (Centre for Ecological and Evolutionary Synthesis, Oslo, [www.cees.com](http://www.cees.com)).

Sekvensene ble sammenlignet med NCBI's database, (National Center for Biotechnology Information) ved bruk av Basic Local Alignment Tool, BLAST. DNA-sekvensene ble redigert ved bruk av dataprogrammet BioEdit og med programmene fra STADEN-pakken. BioEdit er et redigeringsprogram hvor man kan analysere og manipulere nuklein- og proteinsekvenser. STADEN-pakken er en programvare, som brukes til oppbygging av sammenhengende kontiger fra enkeltsekvenser fra sekveringsinstrumenetet. Sekvensene ble oversatt til protein og sammenstilt ved hjelp av ClustalW. ClustalW er et sekvens-sammenstillingsprogram for DNA- eller proteinsekvenser (Thompson *et al.*, 1994; Chenna *et al.*, 2003). Programmet sammenligner et sett med sekvenser parvis alle mot alle, og setter dem sammen slik at likheter og ulikheter tydelig kommer frem.

- BioEdit: [www.mbio.ncsu.edu/BioEdit/bioedit.html](http://www.mbio.ncsu.edu/BioEdit/bioedit.html)
- STADEN: [staden.sourceforge.net](http://staden.sourceforge.net)
- BLAST: [www.ncbi.nlm.nih.gov/BLAST](http://www.ncbi.nlm.nih.gov/BLAST)

## 2.2.2 Fylogenetiske analyser

Fylogenetiske trær ble laget ved bruk av tre forskjellige metoder, Neighbor Joining, Bayesisk interferens og Maximum Likelihood. Neighbor Joining ble utført ved hjelp av programmet MEGA. Bayesisk interferens ble utført ved hjelp av Mr. Bayes og Maximum Likelihood ble utført ved hjelp av PHYML. Mr. Bayes og PHYML ble kjørt med en for optimalisert proteinevolusjonsmodell funnet ved hjelp av Prottest. PhymI og Mr. Bayes er tilgjengelig i UiO Bioportal, ([www.bioportal.uio.no](http://www.bioportal.uio.no)).

### 2.2.2.1 Neighbor-Joining analyser

Neighbor-Joining er en distanse-basert metode som benytter minimums-evolusjons kriterier for å lage fylogenetiske trær (Saitou og Nei, 1987). Metoden begynner med å sette sammen et tre hvor alle sekvensene går gjennom samme punkt ("stjernetre"), så sammenligner den de nærmeste sekvensene mot hverandre, og igjen de som er nærmest dem igjen skrittvis til alle er sammenlignet mot alle. Metoden bruker så avstandsmål til å korrigere for flere treff i samme punkt, og velger en trestruktur som viser summen av de korteste greinene som et estimert tre. Det endelige treet vil ikke nødvendigvis være det mest korrekte treet med den beste trestrukturen fordi det konstruerte treet dannes fra en "skrittvis" algoritme, hvor den samme prosedyren blir gjentatt flere ganger i forsøk på å maksimere utfallet, men det ligner som regel mye på det optimale treet. Neighbor-Joining ble utført ved hjelp av MEGA (Molecular Evolutionary Genetics Analysis), som er et analyseprogram for sammenligning av sekvenser og konstruksjon av fylogenetiske trær. Programmet har funksjoner hvor man kan oppgi beste evolusjonsmodell fra f.eks. Prottest, men ettersom den beste modellen ikke var tilgjengelig i MEGA ble det brukt default substitusjonsmodell.

- MEGA: [www.megasoftware.net](http://www.megasoftware.net)

### 2.2.2.2 Bayesiske analyser

Mr. Bayes (Huelsenbeck *et al.*, 2001; Ronquist og Huelsenbeck, 2003) er et program som konstruerer fylogenetiske trær i et Bayesisk statistisk rammeverk, basert på Markov Chain Monte Carlo simuleringer (Huelsenbeck *et al.*, 2001; Lewis, 2001; Nylander *et al.*, 2004). Programmet beregner trær ut fra alle muligheter i et gitt datasett, og ved å variere parameterne fra gitte evolusjonsmodeller. Først kjører programmet en analyse hvor den forsøker å finne hvilke parameter som gir maksimal sannsynlighet, så lager den mulige trær med verdier rundt de mest sannsynlige parameterne. Avhengig av hvor lenge reaksjonen kjøres blir det laget mange mulige trær, de dårligste trærne forkaster, og de resterende blir satt sammen til et konsensus tre hvor sannsynligheten for grupperingene blir oppført. For å kjøre Mr. Bayes ble optimale parametere fra protest brukt.

Med Mr. Bayes ble det kjørt to analyser, en med covarion og en uten covarion. Covarion tar hensyn til at enkelte sekvenser kan utvikle seg raskere enn andre. Når man analyserer det samme datasettet med to modeller kan man beregne Bayesisk faktor (Kass og Raftery, 1995). Bayesisk faktor er en beregnet sannsynlighet gitt ut fra et datasett, hvor en modell testes mot en annen modell.

- Mr. Bayes: <http://mrbayes.csit.fsu.edu/>

### 2.2.2.3 Maximum Likelihood

Maximum likelihood (Yang, 2000) baserer seg på optimaliserte kriterier for å finne det beste treet av alle mulige trær. Metoden forutsetter at alle evolusjonære endringer er uavhengige av hverandre, for så å prøve ut alle muligheter og beregne hvilke som er mest sannsynlige. I prinsippet skal metoden undersøke alle muligheter og generere et estimat for hvert tre (Whelan og Goldman, 2001). I denne oppgaven består datasettet av 41 sekvenser, dette vil gi omtrent  $3 \times 10^{50}$  mulige trær, og ta en evighet å beregne, så man må begrense antall hypoteser. Her ble det satt til 5.000.000 forsøk. For å utføre maximum likelihood ble det benyttet PHYML (Guindon og Gascuel, 2003), og optimal evolusjonsmodell fra Prottest.

#### 2.2.2.4 Prottest og Mr. Modeltest

Beste modell for evolusjon kan finnes ved bruk av statistiske analyser. Prottest (Abascal *et al.*, 2005) og Mr. Modeltest (Posada og Crandall, 1998) estimerer parametere og beregner hvilken evolusjonsmodell som er best egnet til et gitt sett sekvenser. Prottest beregner utfra en proteinsammenstilling, mens Mr. Modeltest beregner utfra DNA-sammenstilling. Ved å beregne utfra proteinsekvenser får man et mer stabilt resultat ettersom 3.kodon i DNA kan variere uten at aminosyren endres.

Sannsynlighetsmetoder for konstruksjon av fylogenetiske trær avhenger av empiriske modeller for protein evolusjon. Prottest og Mr. Modeltest bruker flere matematiske modeller til å velge hvilken som er den beste modellen for proteinevolusjon til et gitt sett sekvenser. Programmene inkluderer parametere for ulike substitusjonsmodeller, uforandrede aminosyrer, ulikheter mellom sekvensene, og empirisk aminosyre frekvens for modellene.

Modellene som er tilgjengelig i Prottest er Dayhoff (Dayhoff *et al.*, 1978), WAG (Whelan og Goldman, 2001), JTT ((Jones *et al.*, 1992)), mtREV (Adachi og Hasegawa, 1996), VT (Muller og Vingron, 2000), CpREV (Adachi *et al.*, 2000), RtREV (Dimmic *et al.*, 2002), MtMam (Cao *et al.*, 1998) og Blosum62 (Heinkoff og Heinkoff, 1992).

Modellene tilgjengelig i Mr. Modeltest er blant annet JC (Jukes og Cantor, 1969), K2P (Kimura's 2 parameter modell, (Kimura, 1980)), F81 (Felsenstein, 1981), HKY (Hasegawa *et al.*, 1985) og GTR (General Time Reversible, (Lanave *et al.*, 1984)). De ulike modellene beregner alle nukleotidsubstitusjon, fra den enkleste modellen JC der man går utfra alle basene har samme frekvens, til den mest komplekse modellen GTR der alle basene og baseparene har forskjellig frekvens.

En modell for proteinevolusjon indikerer sannsynligheten for forandring i aminosyrene, og modellene antar i utgangspunktet at det er lik sannsynlighet for aminosyreutbytting i alle posisjoner. Programmet beregner derfor også parametere for evolusjon ettersom sannsynligheten for aminosyreutbytting er relatert til strukturen og funksjonen i den aktuelle posisjonen. Informasjon om andel uforandrede aminosyrer i sekvenssammenstillingen vises som +I. Frekvensen av variasjon i et hvert sete kan beskrives ved hjelp av såkalt gamma fordeling. Gamma



fordeling er gitt av to parametere,  $\alpha$  og  $\gamma$ , for kontinuerlig fordeling som representerer summen av tilfeldig gjennomsnittlig eksponentiell fordeling. Når verdien til  $\alpha$  parameteret er mindre enn 1 vil fordelingen ha en L-form, med noen få posisjoner som endres raskt. Når parameteret er over 1 vil fordelingen bli mer normal, med mindre variasjon. Modellene kan også oppgi frekvensen for observerte aminosyre i sammenstilningen av sekvenser, dette oppgis som F.

- Prottest: [darwin.uvigo.es/software/prottest\\_server.html?](http://darwin.uvigo.es/software/prottest_server.html?)
- Mr. Modeltest : <http://darwin.uvigo.es/software/modeltest.html>

### 2.2.2.5 Statistisk støtte for forgreiningsmønster

Til å beregne statistisk støtte for forgreiningsmønsteret for Maximum Likelihood og Neighbor-Joining bruker programmene bootstrapping. Bootstrapping er en metode som estimerer sannsynligheten til trestrukturen (Felsenstein, 1985). Det beregnes en sannsynlighet for hver av forgreiningene i et estimert tre, og oppgir en tallfestet sannsynlighet for hvor ofte denne oppdelingen forekommer ved gjentatte sannsynlighetsberegninger av sekvensene. I denne oppgaven har troverdigheten for trestrukturen ved Maximum Likelihood blitt estimert med 1000 bootstrap forsøk, dvs. at en bootstrapverdi på 1000 tilsvarer 100 % sannsynlighet for den gitte strukturen skal forekomme. For Neighbor-Joining er parameterne oppgitt i prosent. Bootstrap verdier over 95 % regnes for svært god støtte, mens helt ned til 75 % er det moderat støtte.

Mr. Bayes' supportverdiene er basert på posterior sannsynlighet. Posterior sannsynlig av en begivenhet er en betinget sannsynlighet som foreslås når er tatt hensyn til relevante bevis. Maksimal support er 1.00.

Erfaring viser at verdiene fra posterior sannsynlighet relativt sett er høyere enn verdien fra bootstrap analyser, men det er to uavhengige metoder så tallverdiene er ikke direkte sammenlignbare.

## 2.3 Rekombinasjonsanalyser

For analyser av mulig rekombinasjon ble det benyttet studier av informative seter i DNA-sekvensene, SplitsTrees4 og RDP2, med optimaliserte parametere for evolusjon fra Mr. Modeltest.

### 2.3.1.1 SplitsTrees4

Evolusjonære data inneholder ofte forskjellige, og noen ganger motstridene fylogenetiske signaler, noe som ikke alltid vises tydelig i et fylogenetisk tre. SplitsTree4 er et program for analysering og visualisering av evolusjonære data, og kan brukes til å påvise mulige rekombinasjonshendelser mellom sekvenser i en sammenstilling. Dette programmet fremstiller informasjonen som et tre-lignende nettverk (Huson og Bryant, 2006). Er det ingen motstridene informasjon vil nettverket se ut som et tre, mens DNA sekvenser med mulig rekombinasjon viser forgreiningene i et nettverk. Programmet beregner mulige greiner som dannes mellom sekvensene utfra gitt modell og parametere beregnet med Mr. Modeltest.

- SplitsTrees4: <http://www-ab.informatik.uni-tuebingen.de>

### 2.3.1.2 Mosaikkstruktur og informative seter

En metode for å få informasjon om mulig rekombinasjon er å studere informative seter i DNA sekvensene. Dette gjøres ved å sammenstille de aktuelle sekvensene for så å se nærmere på posisjoner hvor minst en til to baser varierer, variasjonene kan si noe om rekombinasjon ved horisontal genoverføring. En mosaikkstruktur forekommer dersom deler av sekvensen stammer fra forskjellige forfedre. En DNA sekvens kan være mosaikk av andre sekvenser som resultat av horisontal genoverføring (Rudi *et al.*, 1998; Boni *et al.*, 2007).

### **2.3.1.3 RDP2**

RDP (Recombination Detection Program) er program for å finne og analysere rekombinasjon i en sammenstilling av DNA-sekvenser (Martin *et al.*, 2005). Programmet bruker ulike modeller til å antyde bruddpunkt i sekvensene som antagelig har oppstått ved horisontal genoverføring. Programmet tester også om funnene er statistiske signifikante. Her er modellene Geneconv, MaxChi og Chimaera benyttet. Testen undersøker tre sekvenser om gangen, og hvilken sekvens som kan ha rekombinert fra to foreldresekvenser.

- RDP2 : [darwin.uvigo.es/rdp/rdp.html](http://darwin.uvigo.es/rdp/rdp.html)

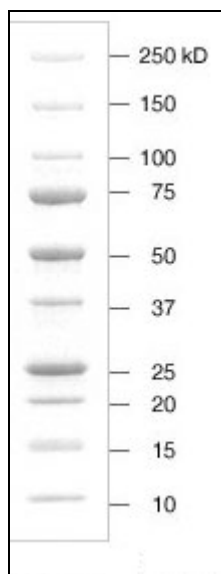
## 2.4 Proteinmetoder

### 2.4.1 SDS-polyakrylamidgelelektroforese

Elektroforese av proteiner blir utført i en polyakrylamidgel. Proteinene behandles med DTT, en reduserende agent ditiotreitol som bryter disulfidbindinger, og SDS, en sterk anionisk detergent (natriumdodecylsulfat), som gir proteinene en negativ ladning. Proteinene blir separert i gelen med hensyn på størrelse. Når en elektrisk spenning blir satt på vil proteinene vandre nedover i gelen, og de minste proteinene vandrer raskest.

Prøvene ble løst i XT Sample buffer fra BioRad, tilsatt DTT og MgCl, og oppvarmet i 5 min ved 95 °C før de ble applisert på gel. Det ble brukt Criterion XT Precast Gel 10 % Bis-Tris fra BioRad, med XT MOPS buffer, gelen ble kjørt ved 200V i 30-60 min. med proteinmarkør ([www.biorad.com](http://www.biorad.com)).

Etter elektroforese kan proteinene bli visualisert ved farging av gelene med Coomassie Brilliant Blue. Gelen avfarges med vann, metanol og eddiksyre.



**Figur 12. Proteinmarkør fra BioRad.**

## 2.5 Heterolog ekspresjon og rensing av enzym

### 2.5.1 Induksjon med IPTG

*E. coli* BL21-celler transformert med ekspresjonsplasmidet ble dyrket i 5 ml LB medium eller NZCYM-medium og inkubert over natten ved 37 °C med risting. Overnatts-kulturen ble videre dyrket opp i 50ml LB-/NZCYM-medium til celletettheten tilsvarte en OD<sub>600</sub> på omtrent 0,7. Proteinekspresjonen ble satt i gang ved å tilsette 1mM IPTG, og det ble tatt prøver for analyse på SDS-PAGE og målt celletetthet hver time.

#### 2.5.1.1 Magic Media

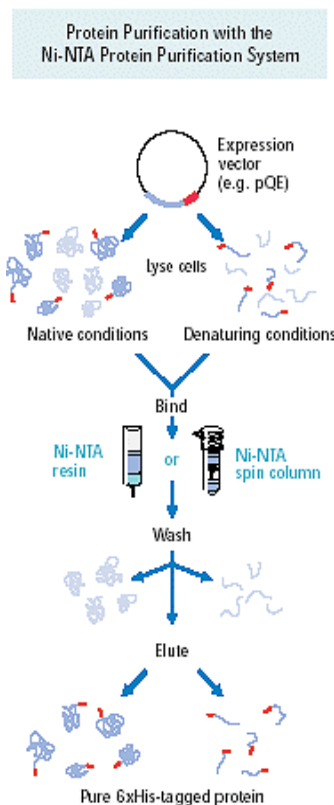
Det ble også brukt MagicMedia™ *E. coli* ekspresjonsmedium fra Invitrogen ([www.invitrogen.com](http://www.invitrogen.com)). Dette er et medium laget for å øke ekspresjonen av rekombinante proteiner i T7-regulerte *E. coli*-ekspresjons systemer uten å måtte følge OD eller tilsette inducere som IPTG. Her ble en forkultur med ekspresjonsplasmidet som er transformert med *E. coli* BL21 celler dyrket opp over natten. Denne kulturen ble overført til MagicMedia slik at forkulturen utgjør 1/20 av totalt volum i MagicMedia.

### 2.5.2 Rensing av His-tag protein

Vektor pET21d innfører en 6xHis-tag i C-enden av proteinet, noe som gjør det enkelt å rense proteinet etter ekspresjon ved affinitetskromatografi. For å gjøre dette ble det brukt Ni-NTA Spin Kit fra Qiagen ([www.qiagen.com](http://www.qiagen.com)). For å rense store mengder protein ble det brukt midi ProPur™ IMAC kit fra NUNC [www.nuncbrand.com](http://www.nuncbrand.com). Ni<sup>2+</sup> inkorporert i kolonnene gjør at His-taggen på proteinet binder seg til kolonnen, og elueres ved bruk av imidazol. Rensing under native forhold bevarer proteinets struktur og funksjon.

For å rense ut protein fra cellene er protokollene brukt slik de er beskrevet for kit'ene. Cellene må først lyses og dette har blitt gjort på to forskjellige måter, enten ved bruk av sonikator eller ved bruk av en French Press.

Cellene som skal renses resuspenderes i lysisbuffer, og inkuberes på is i 30 min før celleveggen ødelegges med sonikator eller French Press.



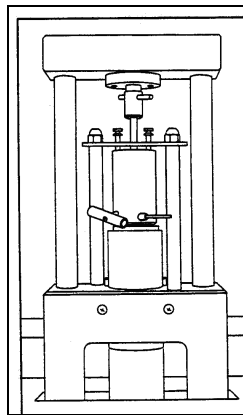
**Figur 13. Proteinrensing med Ni-NTA protein rensingssystem.**

### 2.5.2.1 Sonikator

Sonikator er et instrument som brukes for å lysere cellene, ved å omdanne elektrisk energi til mekanisk energi lages det en intens høy ultralyd. Det ble brukt en UP400S (400 watts, 24 kHz) ultrasonikator fra Hielscher Ultrasonics ([www.hielscher.com](http://www.hielscher.com)). Cellene ble overført til eppendorfrør med ca 50µl i hvert rør, og instrumentet ble innstilt på 0,5 sykluser og amplituden på 60 %. Prøvene ble sonikert på is i 6 x 10 sek, med 5 sek pause mellom.

### 2.5.2.2 French Press

French Press er et apparat som brukes til å ødelegge celleveggen (Vanderheiden *et al.*, 1970). Cellene presses gjennom en sylinder ved høyt trykk og når de kommer igjennom faller trykket til atmosfærisk trykk og celleveggen sprekker. Apparatet brukt er fra Thermo Scientific ([www.thermo.com](http://www.thermo.com)). Det har to celler med ulike volum. I denne oppgaven er den minste cellen brukt. Den tar 3,7ml og et maksimalt trykk på 20.000 psi. Cellene ble presset gjennom sylindren to ganger ved 700atm (1 atm = 14.6956 psi).



**Figur 14. French Press.**

Videre ble lysatet sentrifugert, og supernatanten overført til Ni-NTA kolonnen med lysisbuffer og sentrifugert. Prøver av lysatet og proteinet som ikke har bundet seg til kolonnen ble tatt ut for analyse på SDS-PAGE. Kolonnen ble vasket, og proteinet ble eluert. Bufferne brukt her består av 50mM  $\text{NaH}_2\text{PO}_4$ , 300mM NaCl og imidazol. I lysisbufferen er det 10mM imidazol, i vaskebufferen er det 20mM imidazol, og i elueringsbufferen er det 300mM imidazol. Bundet protein elueres ved at imidazol frigjør det His-taggete proteinet og binder seg til kolonnen.

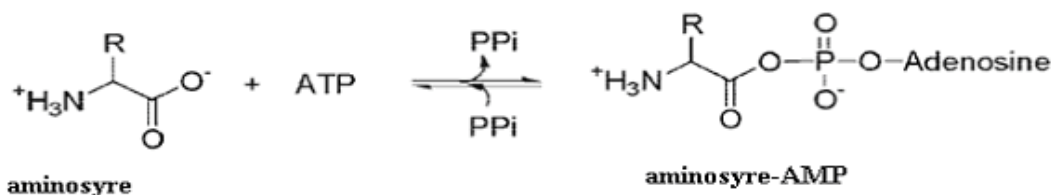
## 2.6 Enzymaktivitet

### 2.6.1 ATP-pyrofosfat utbytting ( $^{32}\text{PPi}/\text{ATP}$ utbytting).

ATP-pyrofosfat-utbytting er en metode for å undersøke aktivitet og spesifisitet av isolerte A-domener, isolerte moduler eller intakte polypeptider fra NRPS systemet (Stachelhaus *et al.*, 1998).

A-domenene gjenkjenner substratet selektivt, og aktiverer aminosyren som acyladenylat ved tilførsel av ATP og frigjøring av PPi. Denne reaksjonen er reversibel, og ved tilførsel av  $^{32}\text{P}$ -merket uorganisk PPi til en reaksjonsmiks vil det dannes  $^{32}\text{P}$ -ATP om aminosyren i reaksjonsmiksen aktiveres av A-domenet. Dannelsen av radioaktivt ATP er proporsjonal med aktiviteten til A-domenet, og kan separeres ved å tilføre aktivt kull til reaksjonen. Kull-bundet radioaktivitet kan måles og gi relative verdier for A-domene aktivitet og selektivitet (Marahiel *et al.*, 1997).

**A**



**B**



Figur 15. ATP/PPi utbyttingsreaksjonen (Lautru og Challis, 2004). Reaksjonen er katalysert av A-domene i NRPS. A viser dannelsen av aminoacyladenylat hvor PPi frigjøres. B viser dannelse av  $^{32}\text{P}$ -ATP.



Reaksjonsløsningen i 100 µl besto av 50mM HEPES med pH 8, 100mM NaCl, 10mM MgCl, 2mM DTT, 1mM EDTA, 0,2mM aminosyre, og det rensede proteinet (Linne og Marahiel, 2004). Reaksjonen ble satt i gang ved å tilføre en startblanding bestående av 4mM ATP, 0.4mM Na<sub>4</sub>PPi og 0.15µCi (16.06Ci/mmol) tetranatrium [<sup>32</sup>P]pyrofosfat.

Løsningen ble inkubert ved 37 °C i 30 min og tilsatt 0.5ml stoppblanding bestående av 1.2 % (vekt/volum) trekull, 0.1M Na<sub>4</sub>PPi og 0.5M perklorsyre.

Videre ble prøvene sentrifugert, slik at kullet samlet seg på bunnen av rørene, og prøvene ble vasket to ganger med 1 ml vann og tilslutt resuspendert i 0.5ml vann.

Tilslutt ble prøvene overført til tellerør, tilsatt Ultima Gold XR, og aktivitet målt i en 1900TR Packard Tri-Carb Liquid Scintillation scintillasjonsteller, et instrument som måler alfa-, beta og gamma aktivitet (Lee og Lipman, 1975; Stachelhaus *et al.*, 1998; Linne og Marahiel, 2004).

Reaksjonen ble testet med alle de vanlige aminosyrene i L-konfigurasjon, og med L-ornitin.

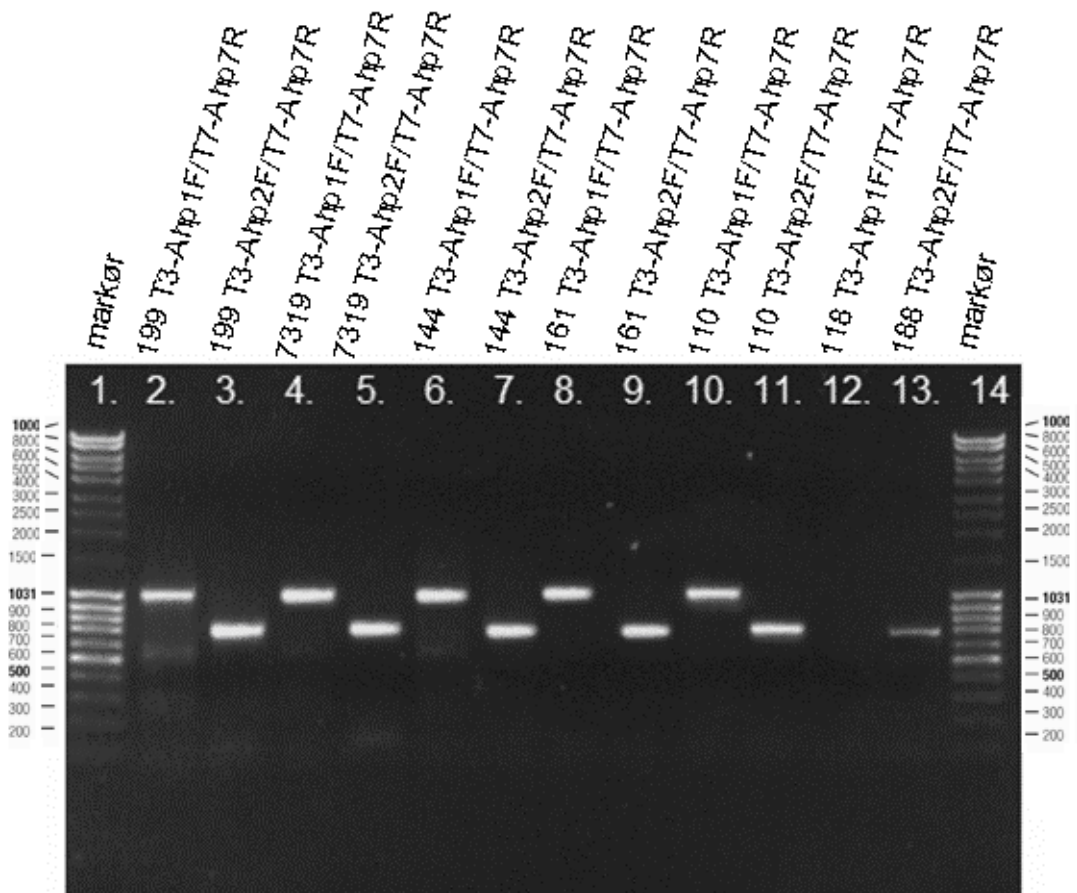
## 3 Resultater

### 3.1 Ahp domenet hos cyanobakterier

#### 3.1.1 PCR analyser

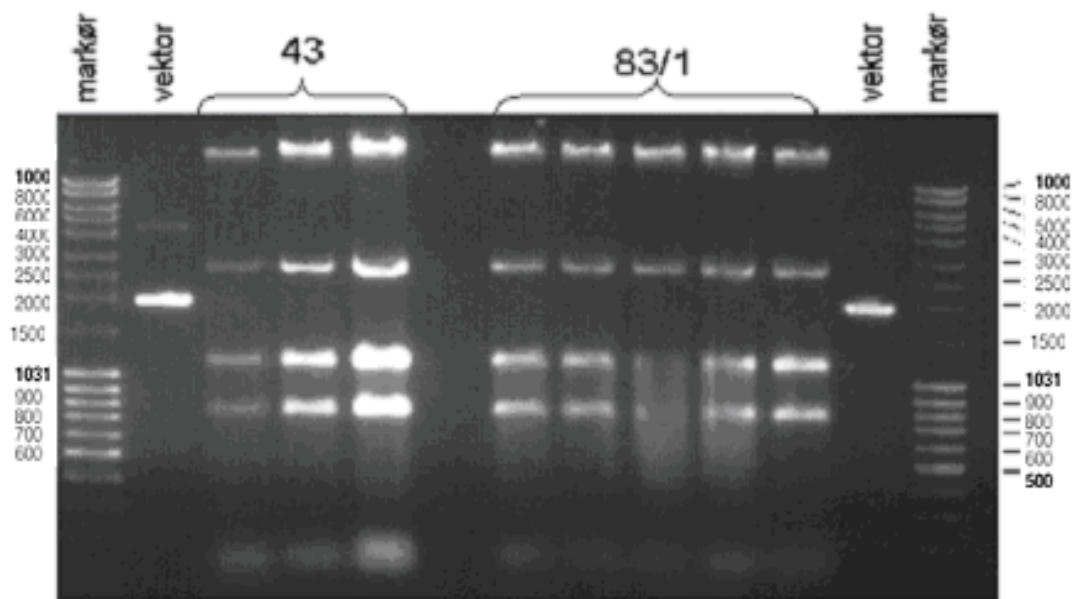
For å studere fordeling og evolusjon av A-domenet som aktiverer Ahp hos cyanobakterier kan man ved hjelp av PCR amplifisere spesifikke sekvenser fra genomisk DNA ved bruk av degenererte primere, som tilsvarer konserverte områder i peptidsyntetasegener. Dette gjør det mulig å identifisere og klonere deler av gener som koder for disse multienzymene. For å studere fordeling og evolusjon av A-domene er det brukt konsensus PCR primere som skiller mellom A-domener som aktiverer Ahp, og A-domener som aktiverer andre aminosyrer.

Ved bruk av primerene i tabell 3 er det kjørt PCR på en rekke ulike cyanobakteriestammer (Tabell 5). PCR-programmet ble optimalisert ved å kjøre PCR med temperaturgradient (55 - 65 °C), og det ble forsøkt med flere ulike primerpar. Det var ønskelig å få et lengst mulig PCR produkt. Derfor ble Ahp1F og Ahp7R foretrukket. Det viste seg at primere med hale ga et mer stabilt resultat. Derfor ble primerpar T3Ahp1F og T7Ahp7R brukt. Av totalt 52 stammer som ble undersøkt var det 30 som ga et PCR produkt i forventet størrelse på 981bp med primerparet T3Ahp1F og T7Ahp7R, og 654bp med T3Ahp2F og T7Ahp7R (Figur 16).



**Figur 16. PCR på ulike bakteriestammer med primer par T3-Ahp1F/T7-Ahp7R (bånd på 981 bp), og T3-Ahp2F/T7-Ahp7R (bånd på 654 bp). 1 og 2 er NC 199, 3 og 4 er PCC 7319, 7 og 8 er NC 144, 10 og 11 er NC 110, 12 og 13 er NC 118.**

PCR produktene ble renset fra agarosegel, ligert inn i pGEM-T Easy vektoren og transformert med kompetente *E. coli* DH5  $\alpha$  celler. Fem positive kolonier fra hver stamme ble undersøkt ved kvikksjekk (Figur 17), som kontroll på at vektoren har tatt opp et plasmid med innskudd. Positive kolonier ble plukket fra LB-skåler, og plasmidet renset med Wizard miniprep som beskrevet i materiell og metoder, og levert til sekvensering ved ABI laboratoriet til CEES, med sekvenseringsprimere pUC/M13 og SP6 (tabell 9).



Figur 17. Kvikksjekk av positive kolonier, NC 43 og NC 83/1

Tabell 5 viser en oversikt over hvilke stammer som har blitt undersøkt i denne oppgaven, og hvilke som har gitt PCR-produkt i forventet størrelse.

Tabell 5 Bakteriestammer

NR	<i>Slekt/art</i>	Opprinnelse	PCR-produkt	BLAST-søk med NRPS-treff
NC 43	<i>Microcystis aeruginosa</i>	Lake Mendota, Wisconsin, USA, 1948	+	+
NC 57	<i>Microcystis aeruginosa</i>	Frøylandsvannet, Rogaland, Norge, 1978	+	
NC 83/1	<i>Anabaena lemmermannii</i>	Edlandsvatn, Rogaland, Norge, 1981	+	+
NC 98	<i>Planktothrix profilica</i>	Steinsfjorden, Buskerud, Norge, 1982	+	+
NC 110	<i>Phormidium formosum</i>	Saka, Kanto, Japan, 1977	+	
NC 118/2	<i>Microcystis sp.</i>	Gjersjøen, Akershus, Norge, 1983	+	
NC 144	<i>Microcystis cf. flos-aquae</i>	Borrevann, Vestfold, Norge, 1984	+	+
NC 161/1	<i>Microcystis botrys</i>	Mosvann, Rogaland, Norge, 1985	+	+
NC 172/5	<i>Microcystis cf. wesenbergii</i>	Arresø, Danmark, 1985	+	+
NC 199	<i>Merismopedia angularis</i>	Sørneset, Spitsbergen, Svalbard, 1984	+	
NC 203	<i>Phormidium sp.</i>	Svalbard, 1984	+	
NC 228/1	<i>Microcystis aeruginosa</i>	Akersvannet, Vestfold, Norge, 1987	+	+
NC 276/6	<i>Pseudoanabaena limnetica</i>	Mälaren, Görveln, Sverige, 1990	+	
NC 278	<i>Planktotrix mougenotii</i>	Steinsfjorden, Norge, 1989	+	
NC 306	<i>Nostoc sp.</i>	Norge, 1991	+	

## Resultater

NC 372	<i>Aphanizomenon sp.</i>	Balatonsjøen, Ungarn, 1996	+	
NC 17	<i>Anabaena cf. Cicalialis</i>	Rødenessjøen, Østfold, Norge, 1970	+	
NC 358	<i>Anabaena cf. Spiroides</i>	Balatonsjøen, Ungarn, 1996	+	
NC 82	<i>Anabaena circinalis</i>	Steinsfjorden, Buskerud, Norge, 1980		
NC 418	<i>Anabaena crassa</i>	Storvatnet, Møre og Romsdal, Norge, 2000		
NC 74	<i>Anabaena crassa</i>	Østensjøvannet, Oslo, Norge, 1980		
NC 147	<i>Anabaena daNiva</i>	Langen, Akershus, Norge, 1984		
NC 138	<i>Anabaena flos-aquae</i>	Hastings Lake, Alberta, Canada, 1977		
NC 139	<i>Anabaena flos-aquae</i>	Beaverhill Lake, Alberta, Canada, 1972	+	+
NC 223	<i>Anabaena flos-aquae</i>	Ukjent, Ungarn		
NC 267/4	<i>Anabaena flos-aquae</i>	Flammestadtjønni, Hordaland, Norge, 1990	+	
NC 269/6	<i>Anabaena flos-aquae</i>	Frøylandsvannet, Rogaland, Norge, 1990	+	+
NC 419/2	<i>Anabaena inaequalis</i>	Storvatnet, Møre og Romsdal, Norge, 2000	+	+
NC 270/1	<i>Anabaena lemmermannii</i>	Arefjordvatnet, Hordaland, Norge, 1990		
NC 32	<i>Anabaena lemmermannii</i>	Burton Lake, Saskatchewan, Canada, 1961		
NC 281/1	<i>Anabaena lemmermannii</i> var. <i>Lemmermannii</i>	Storavatn, Hordaland, Norge, 1990		
NC 298	<i>Anabaena lemmermannii</i> var. <i>Lemmermannii</i>	Storavatn, Hordaland, Norge, 1991		
NC 266/1	<i>Anabaena lemmermannii</i> var. <i>minor</i>	Bergesvatn, Hordaland, Norge, 1990		
NC 426	<i>Anabaena lemmermannii</i> var. <i>Minor</i>	Storvatn, Hordaland, Norge, 2000		
NC 282	<i>Anabaena oscillariodes</i>	Nilen, Egypt, 1987		
NC 66	<i>Anabaena planctoNiva</i>	Langsævatnet, Aust-Agder, Norge, 1979		
NC 416	<i>Anabaena solitaria</i>	Bergesvatnet, Møre og Romsdal, Norge, 1999		
NC 233	<i>Anabaena sp.</i>	Ukjent, Rusisk reservoar		
NC 246	<i>Anabaena sp.</i>	Ukjent, ca 1971		
NC 44	<i>Anabaena sp.</i>	Lake Mendota, Wisconsin, USA, 1948		
NC 140	<i>Microcystis aeruginosa</i>	Bendig's pond, Saskatchewan, Canada, 1975		
NC 160/2	<i>Microcystis aeruginosa</i>	Akersvannet, Vestfold, Norge, 1985		
NC 162/3	<i>Microcystis aeruginosa</i>	Arresø, Danmark, 1985		
NC 22	<i>Microcystis aeruginosa</i>	Lake Mendota, Wisconsin, USA, 1948	+	+
NC 465	<i>Microcystis aeruginosa</i>	Victoria, Murchison Bay, Uganda, 2002		
NC 171/10	<i>Microcystis botrys</i>	Arresø, Danmark, 1985	+	+
NC 357	<i>Microcystis botrys</i>	Zala, Ungarn, 1996	+	+
NC 432	<i>Microcystis novacekii</i>	Viktoriasjøen, Murchison Bay, Uganda, 2000	+	
NC 496	<i>Microcystis sp.</i>	Queen Elizabeth Channel, Uganda	+	
NC 547	<i>Microcystis sp.</i>	Pilsner Damm, Kenya, 2002	+	
PCC 7806	<i>Microcystis sp.</i>	Braakman, Nederland, 1972	+	+
PCC 7319	<i>Pleurocapsa</i>	Arizona Marine Station, Puerto Penasco, Mexico, 1971	+	
NC 116	<i>Planktothrix</i>	Årungen, Akershus, Norge, 1983 (Rounge <i>et al.</i> , 2007)		
NIES 205	<i>Planktothrix agardhii</i>	Komarek, Japan (Rounge <i>et al.</i> , 2007)		
90	<i>Anabaena</i>	Vesijärvi, Finland (Rouhiainen <i>et al.</i> , 2000)		

NC, NIVA-CYA, Norsk institutt for vannforskning

PCC = Pasteur Culture Collection

NIES = National Institute for Environmental Studies, Japan

### 3.1.2 Sekvensanalyser

Sekvensene ble undersøkt ved BLAST-søk, og alle som ikke ga NRPS-treff ble eliminert. Av totalt 30 stammer som ga PCR produkt som ble klonet og sekvensert var det 16 som ikke ga NRPS treff. De resterende 14 sekvensene ble redigert ved bruk av STADEN-pakken og BioEdit. Sekvensene ble sammenstilt ved bruk av Clustal W. Sekvensene ble også sammenstilt med publiserte sekvenser fra anabaenopeptilid syntetase operonet (adp) fra *Anabaena* 90 (Rouhiainen *et al.*, 2000), ocillapeptin syntetase operon (oci) fra *Planktothrix* NC 116 (Rounge *et al.*, 2007) og *Planktothrix* NIES 205 (Trine B. Rounge).

#### 3.1.2.1 Konserverte områder

Sekvenssammenstillingen ble undersøkt for å kontrollere at det var A-domener ved å sammenligne sekvensene mot de konserverte områdene i A-domenet (Tabell 2). De konserverte områdene A3-A7 ble funnet. Sekvenssammenstillingen er vist med markering av konserverte områder i Figur 36.

#### 3.1.2.2 Signatursekvenser

Sekvensene ble sammenstilt med aminosyresekvensen til GrsA-fenylalanin A-domenet (Stachelhaus *et al.*, 1999), og signaturen til antatte aminosyrebindingsete posisjoner ble funnet. Signatursekvensen kan forutsi hvilket substrat A-domenet aktiverer. Det er samme aminosyre i de gitte posisjonene til alle Ahp sekvensene som er funnet her, signatursekvensen er DVENAGVVT (Tabell 6, Figur 36).

**Tabell 6 Sammenligning av signatursekvenser til A-domener som aktiverer Ahp, med publiserte signatursekvenser til A-domener som aktiverer fenylalanin, ornitin, glutamin, isoleucin og glutamat (Stachelhaus *et al.*, 1999; Rouhiainen *et al.*, 2000; Tooming-Klunderud *et al.*, 2007)**

	<b>235</b>	<b>236</b>	<b>239</b>	<b>278</b>	<b>299</b>	<b>301</b>	<b>322</b>	<b>330</b>	<b>331</b>	<b>517</b>
Fenylalanin	D	A	W	T	I	A	A	V	C	K
<b>Ahp</b>	<b>D</b>	<b>V</b>	<b>E</b>	<b>N</b>	<b>A</b>	<b>G</b>	<b>V</b>	<b>V</b>	<b>T</b>	<b>?</b>
<b>mcnE</b>	<b>D</b>	<b>A</b>	<b>Q</b>	<b>D</b>	<b>L</b>	<b>G</b>	<b>V</b>	<b>V</b>	<b>D</b>	<b>K</b>
Ornitin	D	V	G	E	I	G	S	I	D	K
Glutamin	D	A	Q	D	L	G	V	V	D	K
Isoleucin	D	A	F	F	L	G	V	V	Y/F	K
Glutamat	D	A	W/K	H/D	F/L	G	G/V	V	D	K

### 3.1.3 Fylogenetiske analyser

Fylogenetiske analyser ble utført på sekvensene sekvensert her, og med publiserte sekvenser fra A-domener fra *Anabaena* (Rouhiainen *et al.*, 2000), *Planktothrix* (Rounge *et al.*, 2007) og *Microcystis* (Tooming-Klunderud *et al.*, 2007).

Fylogenetiske trær ble konstruert ved brukt av Neighbor-Joining metoden i MEGA, maximum likelihood i Phym, og med Bayesiske analyser Mr. Bayes. For Neighbor-Joining metoden ble det ikke definert noen substitusjonsmodell, men for å finne beste evolusjonsmodell for proteinsekvensene til Phym og Mr. Bayes ble Prottest brukt.

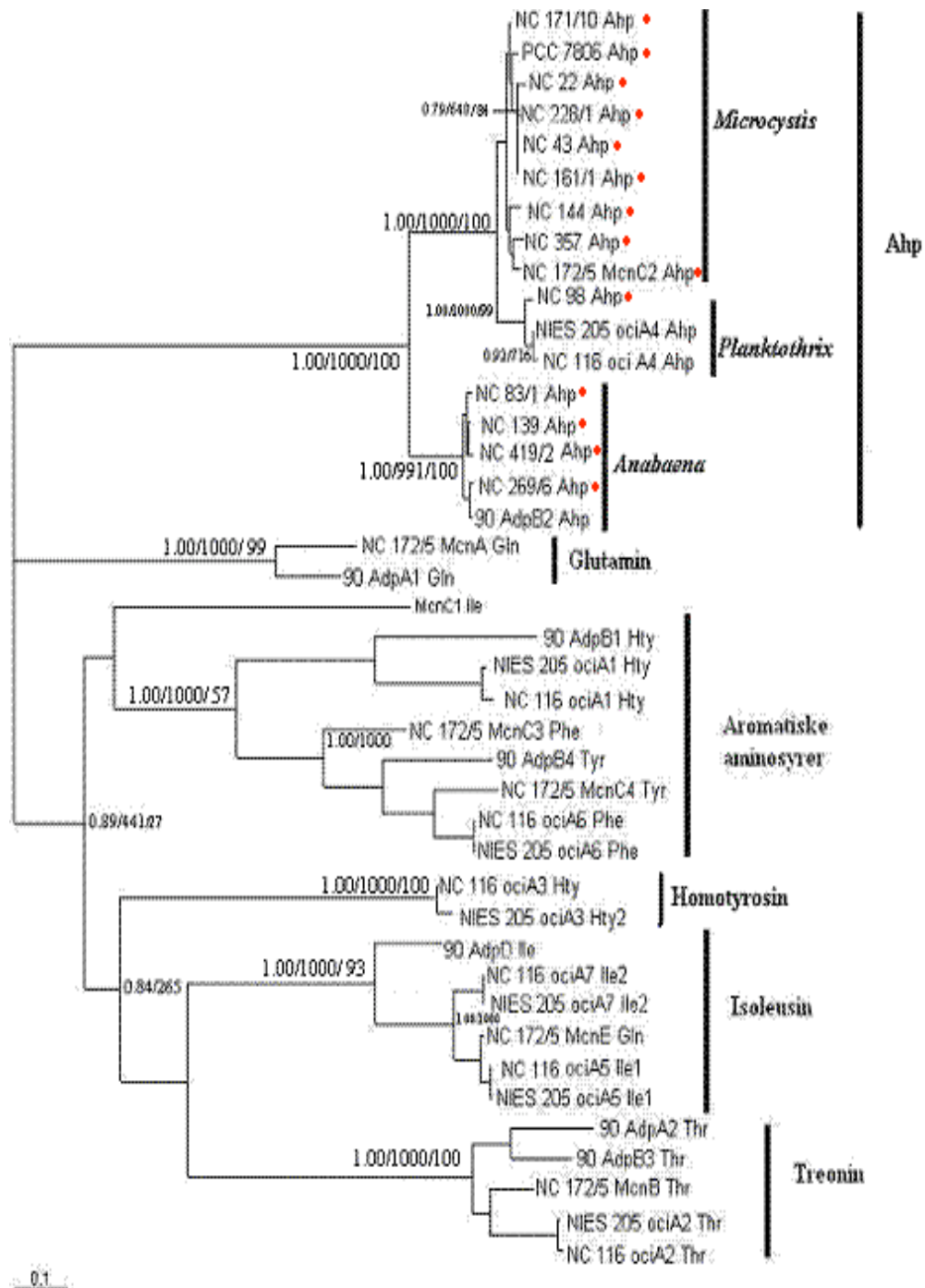
Prottest oppga at beste modell var WAG+I +G +F. Modellen tar derfor også med gamma fordeling +G=0,15, men den tar ikke med uforandrede aminosyrer alene +I=0,00. Sammen med gamma fordelingen beregnes +I+G=0,85 og aminosyrefrekvensen +F = 1,00.

Med Mr. Bayes ble det kjørt to analyser en med covarion og en uten covarion. Mr. Bayes beregner en gjennomsnittlig faktor for hver av de to modellene, uten covarion

er den -7230,25, og med covarion er den -7227,65. Bayesisk-faktor beregnes ved å ta forskjellen mellom de to verdiene og multiplisere med 2, dette gir 5,38. Dette gir ingen signifikant forskjell mellom de to modellene. For at det skal være en forskjell på modellene må Bayesisk-faktor være over 10 (Kass og Raftery, 1995). De to trærne ble helt like, med kun minimale forskjeller i bootstrapverdiene.

De ulike metodene Neighbor-Joining, Maximum Likelihood, og Bayesisk analyse ga samme trestruktur så et tre vises her med statistisk sannsynlighet fra de ulike metodene angitt på hver node (Figur 18). For Neighbor-Joining og Maximum likelihood er statistiskstøtte beregnet ved bootstrapanalyser, for Bayesisk er det beregnet utfra posterior sannsynlighet.





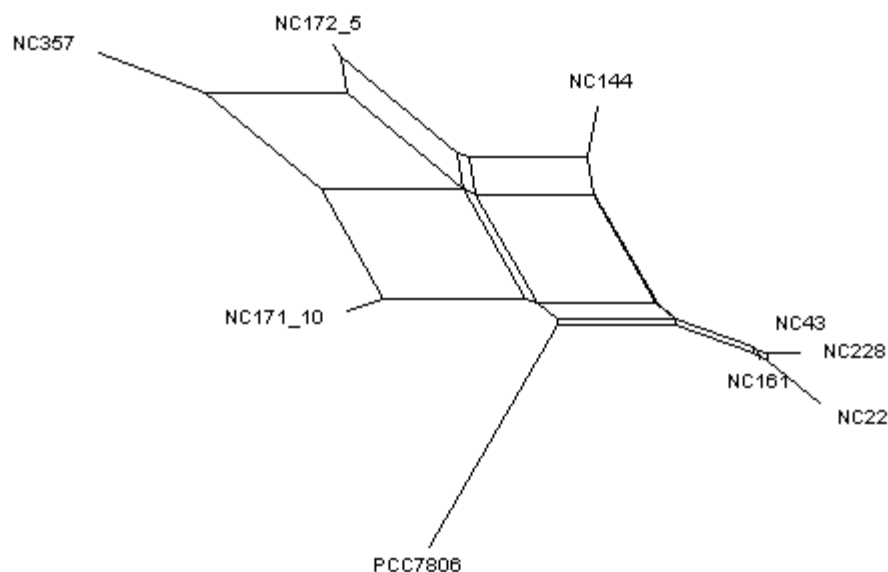
**Figur 18.** Fylogenetisk analyse av A-domener fra ulike cyanopeptolinsyntetaser. NC 172/5 Mcn er cyanopeptolinsyntetaser fra *Microcystis* (Tooming-Klunderud *et al.*, 2007). 90 Adp er anabaenopeptilidsyntetase fra *Anabaena* 90 (Rouhiainen *et al.*, 2000), NIES 205 oci og NC 116 oci er ocillapeptin syntetase fra *Planktothrix* (Rounge *et al.*, 2007). F. eks er NC 172/5 McnA Gln er et A-domene som aktiverer av glutamin fra NC 172/5 Mcn. Sannsynlighetene for hver node vises fra Mr. Bayes/PhyML/NJ, 1.00/1000/100 vil si at alle programmene støtter noden med 100 %. Sekvensene med røde stjerner er de jeg har funnet i dette arbeidet.

### 3.1.4 Rekombinasjons analyser

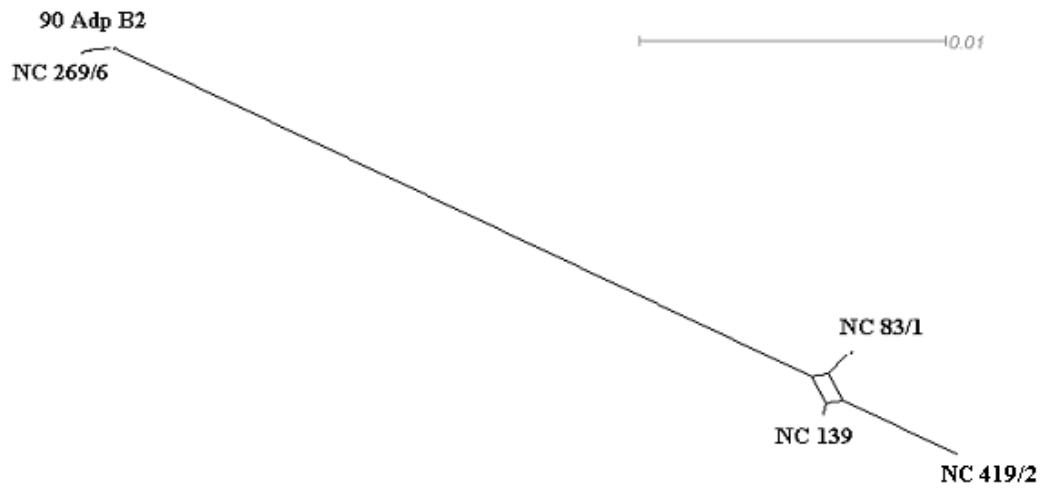
#### 3.1.4.1 SplitsTrees

Evolusjonært slektskap blir ofte fremstilt med fylogenetiske trær. Ved evolusjon hvor det kan ha forekommet horisontal genoverføring eller rekombinasjon vil ikke et vanlig fylogenetisk tre vise dette. SplitsTrees gir informasjon om rekombinasjon mellom stammene som gjerne er klustrer nært sammen i et fylogenetisk tre (Huson og Bryant, 2006). SplitsTrees4 lager et fylogenetisk nettverk over mulig rekombinasjon mellom gitte sekvenser. Her er *Microcystis*, *Anabaena* og *Planktothrix* inndelt hver for seg. Analysene er kjørt med DNA-sekvenser og Mr. Modeltest er brukt for å finne beste modell og optimale parametere. Beste modell er GTR+I (General Time Reversible) modell, der basefrekvensene er A=0,2785, C=0,2076, G=0,220 og T=0,2936.

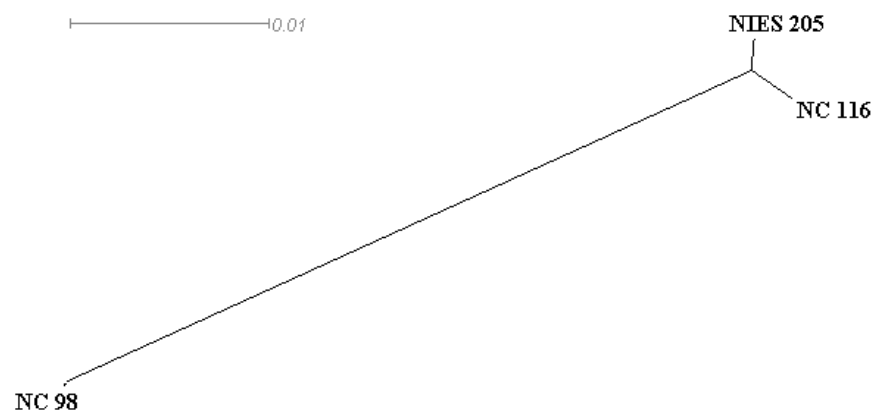
For å måle likheten beregner programmet en parvis homoplasi indeks (Phi) (Bruen *et al.*, 2006), en verdi for sannsynlig rekombinasjon. For *Microcystis* (Figur 19) fant phi-testen det sannsynlig at rekombinasjon kan forekomme,  $p=5,56 \times 10^{-6}$ . For *Anabaena* (Figur 20) og *Planktothrix* (Figur 21) fant phi-testen ingen statistisk signifikans for rekombinasjon,  $p=1,0$ .



**Figur 19. SplitsTrees analyse av A-domener fra *Microcystis*.**  
NC er stammer fra NIVA, PCC er fra Pasteurinstituttet i Frankrike. Figuren viser at det dannes et nettverk som viser at det er rekombinasjon mellom stammene



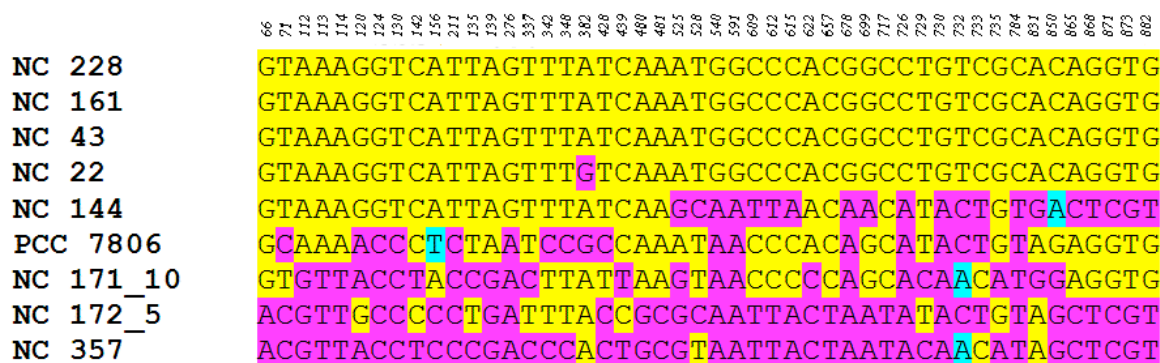
**Figur 20. SplitsTrees analyse av A-domener fra *Anabaena*.**  
 NC er stammer fra NIVA, 90 Adp B2 er *Anabaena* 90 publisert av Rouhiainen (2000).



**Figur 21. SplitsTrees analyse av A-domener fra *Planktothrix*.**  
 NIES 205 og NC 116 er sekvenser fra Trine B. Rounge (2007)

### 3.1.4.2 Mosaikk struktur

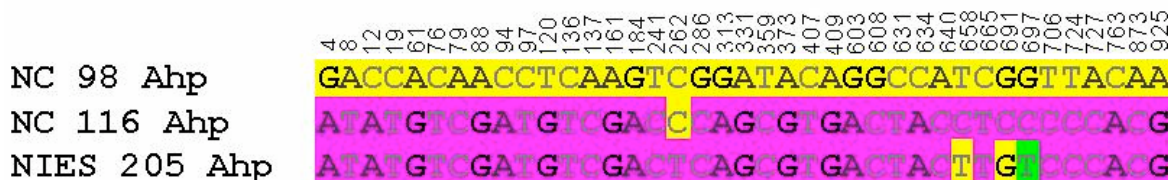
DNA-sekvenser kan fremstå som mosaikker av andre sekvenser som resultat av rekombinasjon ved horisontal genoverføring. Her er DNA-sekvensene fra de ulike stammene sammenstilt og sammenlignet for å se om det er noe mønster, som kan gi informasjon om rekombinasjon. Informative seter hvor minst en base varierer og hver variant finnes i minst to stammer er markert på figurene for *Microcystis* (Figur 22), *Anabaena* (Figur 23) og *Planktothrix* (Figur 24).



Figur 22. Mosaikkstruktur til informative seter hos *Microcystis*. Farge kodene er valgt for å illustrere sekvenser vist med informative seter som "hører sammen", det vil si at de har samme fylogenetiske affinitet.



Figur 23. Mosaikkstruktur til informative seter hos *Anabaena*. Fargekoder som figur 22.



Figur 24. Mosaikkstruktur til informative seter hos *Planktothrix*. Samme fargekoder som figur 22.

### 3.1.4.3 RDP2 (Recombination Detection Program)

RDP2 ble brukt for å se nærmere på mulig rekombinasjon (Martin *et al.*, 2005). Programmet identifiserte sekvenser som kan være rekombinante, hvor rekombinasjonen skjer, hvilken sekvens som recombines. Programmet beregner sannsynligheten for rekombinasjonen, vist i tabellen med p. Geneconv, Maxchi og Chimaera ga her informasjon om rekombinasjon hos *Microcystis*, og resultatet *Microcystis* er vist i tabellen under (Tabell 7). For *Anabaena* og *Planktothrix* ble det ikke funnet noe informasjon på rekombinasjon.

**Tabell 7 Mulig rekombinasjon hos *Microcystis* funnet med RDP2**

Stammer involvert	Geneconv	Maxchi	Chimaera
NC 144 av NC 357 og NC 161	1-521 bp p=2,53e-4		
NC 144 av NC 357 og NC 43	1-524 bp p=2,53e-4	71-481 bp p=7,85e-6	459-882 bp p=2,72e-6
NC 144 av NC 22 og NC 357		629-882 bp p=2,88e-8	
NC 144 av NC 171/10 og NC 22		35-327 bp p=1,33e-4	
NC 144 av NC 22 og NC 172/5		525-882 bp p=1,19e-4	
NC 144 av PCC 7805 og NC 161		108-435 bp p=9,05e-7	330-870 bp p=7,48e-6
NC 144 av NC 228 og NC 357		540-882 bp p=5,8e-6	
NC 144 av NC 172/5 og NC43			66-481 bp p=2,2e-5

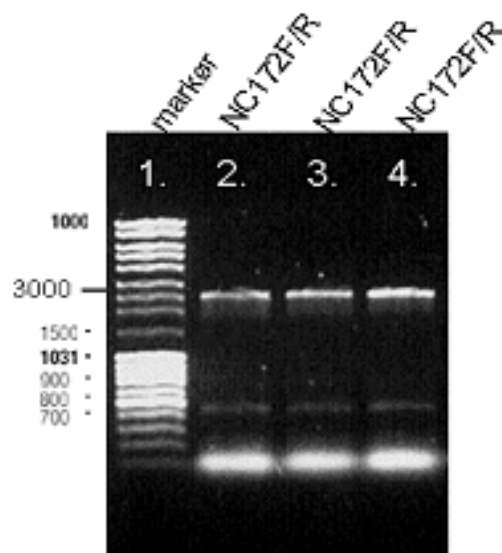
## 3.2 Ahp domeneaktivitet

### 3.2.1 Ekspresjon av Ahp-domenet fra stamme 172

Det var ønskelig å finne ut hvilket molekyl som er forløperen for Ahp-delen som finnes i cyanopeptoliner. En metode for å finne ut dette er å uttrykke A-domenet fra Ahp-modulen til NC 172/5 *Microcystis cf. wesenbergii* og bestemme hvilke aminosyrer dette domenet aktiverer. Det er kjent at NC 172/5 *Microcystis cf. wesenbergii* produserer cyanopeptolin (Tooming-Klunderud *et al.*, 2007).

### 3.2.2 PCR

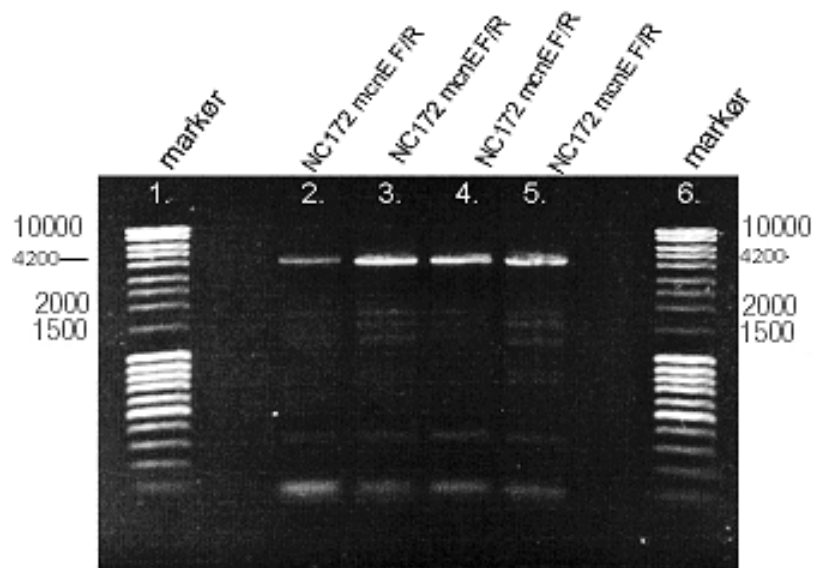
For å gjøre dette er det brukt primere spesielt laget for kloning av domener i ekspresjonsvektor. PCR ble kjørt med primerpar Ahp\_172F/R (tabell 1), og det samme programmet som tidligere med Phusion polymerase (se seksjon 2.1.3 om PCR). Dette ga et PCR-produkt i forventet størrelse på 2600bp (figur 25).



Figur 25

**Figur 25. PCR av Ahp med primerpar Ahp\_172F/R (bånd på 2800)**

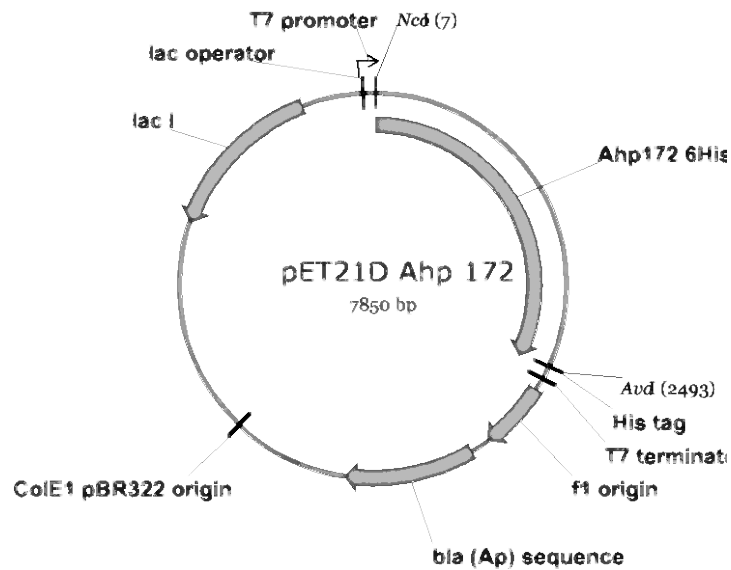
Som kontrollprotein til Ahp domenet er mcnE domenet fra NC 172/5 brukt. Det ble valgt fordi det også her er ukjent hvilken aminosyre det aktiverer. PCR ble utført med primerpar mcnE-F/R under de samme betingelsene som tidligere (se seksjon 2.1.3 om PCR), dette ga et PCR-produkt i forventet størrelse på ca 4200bp (Figur 26).



Figur 26. PCR av mcnE med primerpar mcnE-F/R (bånd på 4200).

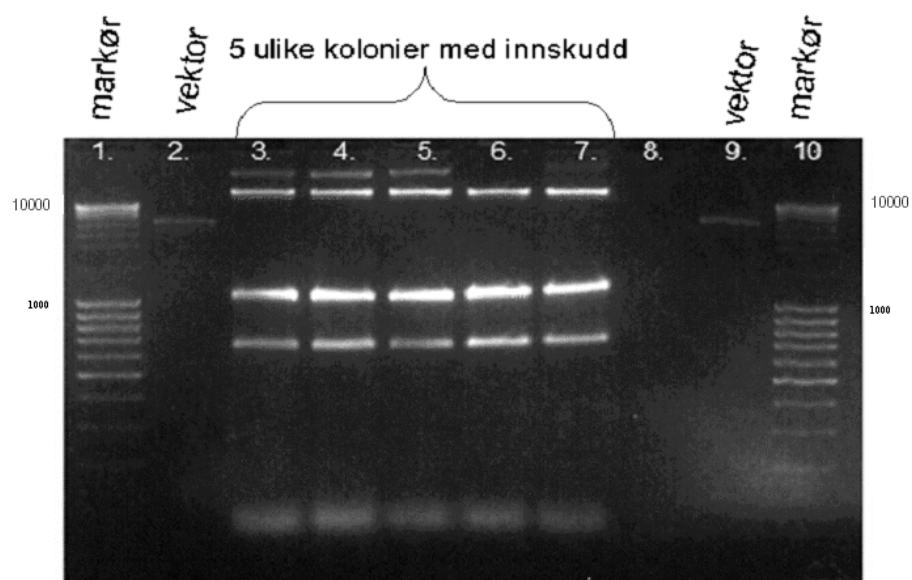
### 3.2.3 Subkloning i pET21d-vektor

PCR-produktene ble rensset fra agarosegel, og kuttet med restriksjonsenzymene NcoI og Aval. Kuttet PCR produkt ble subklonet i en pET-vektor (figur 27). pET21d inneholder en T7 RNA polymerasepromotor, noe som gjør den godt egnet for protein ekspresjon. Videre ble pET21d vektor og PCR-produktene transformert med *E. coli* DH5α.



Figur 27. pET21d Ahp plasmid (Tom Kristensen).

Positive kolonier ble undersøkt med kvikksjekk (kun *mcnE* er vist i Figur 28), isolert og sendt til sekvensering med sekvenseringsprimere T7 og T7term (tabell 9) for å bekrefte at produktet var riktig.

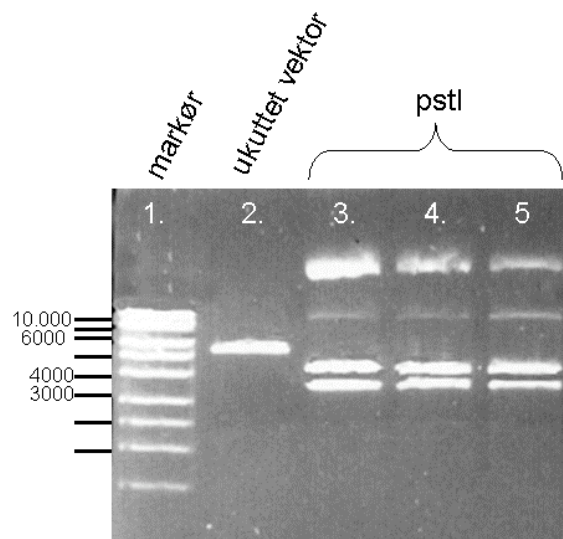


Figur 28. kvikksjekk av *mcnE* fra NC 172/5.

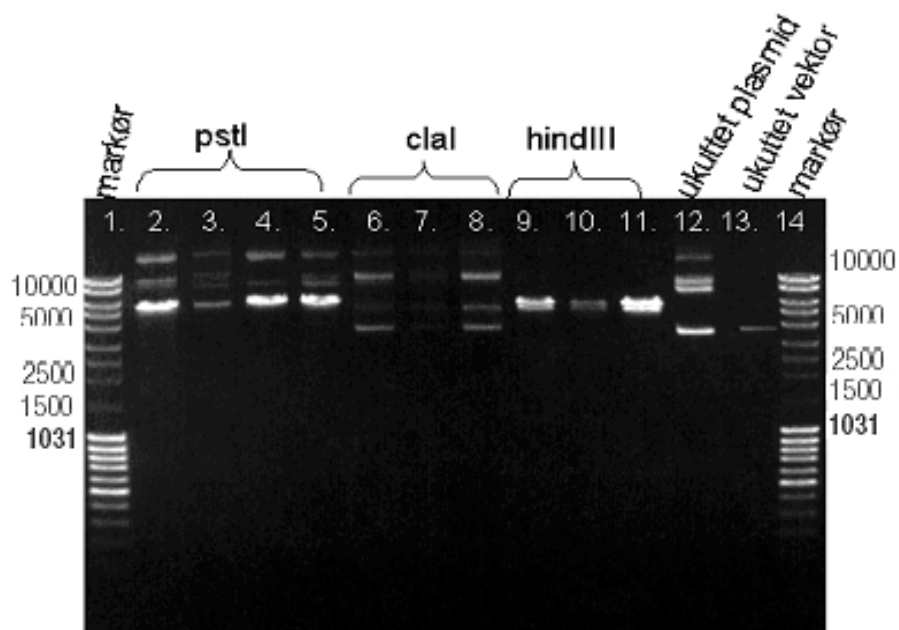


### 3.2.4 Transformasjon av *E. coli* BL21

pET21d Ahp 172 og pET21d mcnE 172 ble transformert inn i *E. coli* BL21 celler. For å være sikker på at det var riktig plasmid som var transformert ble klonene testkuttet med restriksjonsenzym. Ahp 172 ble testkuttet med pstI (Figur 29), mcnE 172 ble testkuttet med pstI, claI og hindIII (Figur 30). mcnE ble også fullstendig sekvensert i begge retninger med 6 sekvenseringsprimere (Tabell 9).



Figur 29. Ahp testkuttet med pstI, ukuttet vektor med bånd på 5400, Ahp med to bånd; 4200 og 3600, ukuttet vektor vises med bånd på 5400bp

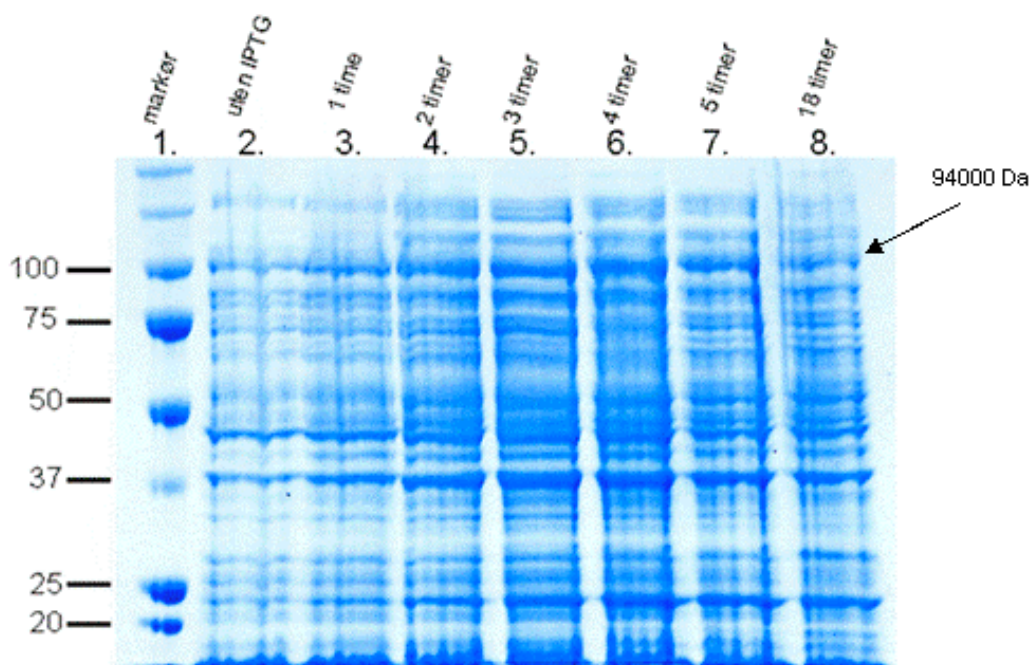


Figur 30. mcnE testkuttet med pstI, claI, og hindIII. pstI gir bånd på 5000bp, claI gir bånd på 3000bp og 5000bp, hindIII gir bånd på 8000bp. Ukuttet plasmid med bånd på 4200bp, ukuttet vektor med bånd på 5400bp.

### 3.2.5 Proteinekspresjon

*E. coli* BL21 celler er godt egnet for proteinekspresjon, her er det brukt *E. coli* BL21(DE3)LysS. LysS plasmidet koder for T7 lysozym, som er en naturlig inhibitor for T7 RNA polymerase, og derfor hindrer genet i å bli transkribert. Ved induksjon med IPTG vil det produseres mye større mengder T7 RNA polymerase enn det T7 lysozymet greier å inhibere.

*E. coli* BL21 celler med pET21d Ahp ble dyrket med NZCYM medium med ampicillin og kloramfenikol og tilsatt IPTG for proteinekspresjon. *E. coli* BL21(DE3)LysS har kloramfenikolresistens, og pET21d har ampicillinresistens. Prøver ble tatt ut hver time før og etter at IPTG var tilsatt for analyse på SDS-PAGE. Figur 31 viser at allerede en time etter induksjon med IPTG blir proteinet synlig på gelen, og etter 4 timer er båndet ganske tydelig, noe som indikerer relativt høy konsentrasjon av proteinet.



**Figur 31. Tidskurve for Ahp fra SDS-PAGE analyse, forventet molekylvekt 94000.**  
Nr 2 er før IPTG er tilsatt, og videre er det tatt prøver for hver time.

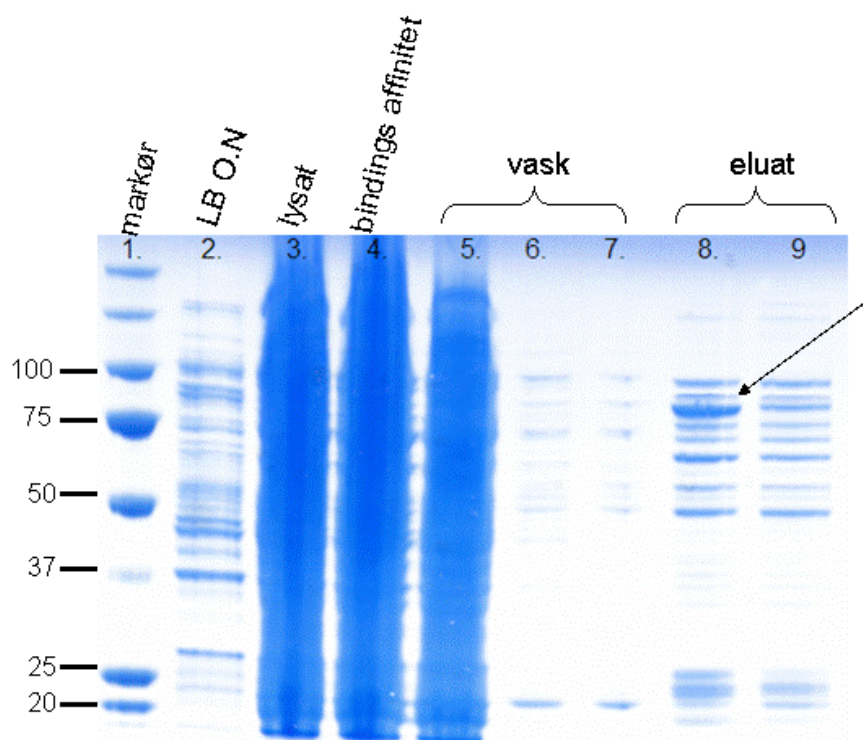
Det var ønskelig å få større proteinekspresjon, så det ble gjort forsøk med MagicMedia™ *E. coli* ekspresjonsmedium fra Invitrogen. Dette er et medium laget for å øke proteinekspresjonen av rekombinante proteiner i T7-regulert *E. coli* ekspresjons systemer uten å overvåke O.D. eller tilføre indukere som IPTG.

### 3.2.6 Rensing av His-tag protein

pET21d innfører en His-tag i C-enden av proteinet, noe som gjør det enkelt å rense proteinet etter ekspresjonen ved affinitetskromatografi på en nikkel kolonne.

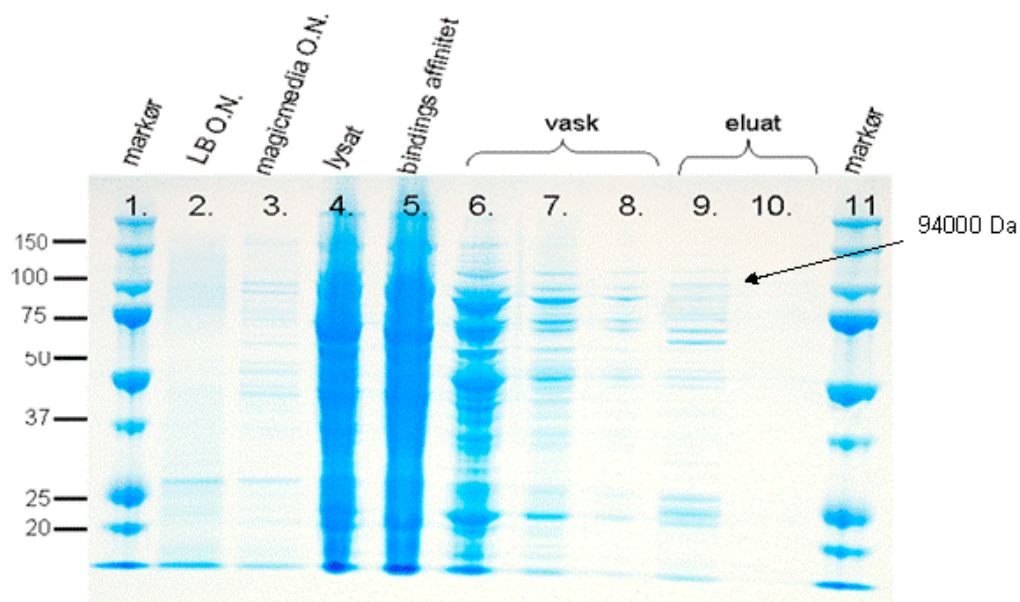
Prøvene ble renset under native forhold ved bruk av Ni-NTA Spin kit fra Qiagen og større mengder prøve ble renset ved bruk av midi kolonner ProPur™ IMAC kit fra NUNC. Prøver fra hvert skritt i rensingen ble samlet og analysert på SDS-PAGE. Forventet molekylvekt for Ahp172 er 94.000 Da, og for mcnE172 er det 140.000.

Dette ble gjort først for pET21dAhp konstruktet. Figur 34 viser et tydelig bånd i forventet størrelse (Figur 32). Eluatet ble lagret ved -20 °C.



**Figur 32. SDS-PAGE analyse for Ahp domene renset ved affinitetskromatografi. Forventet størrelse på proteinet 94000Da.**

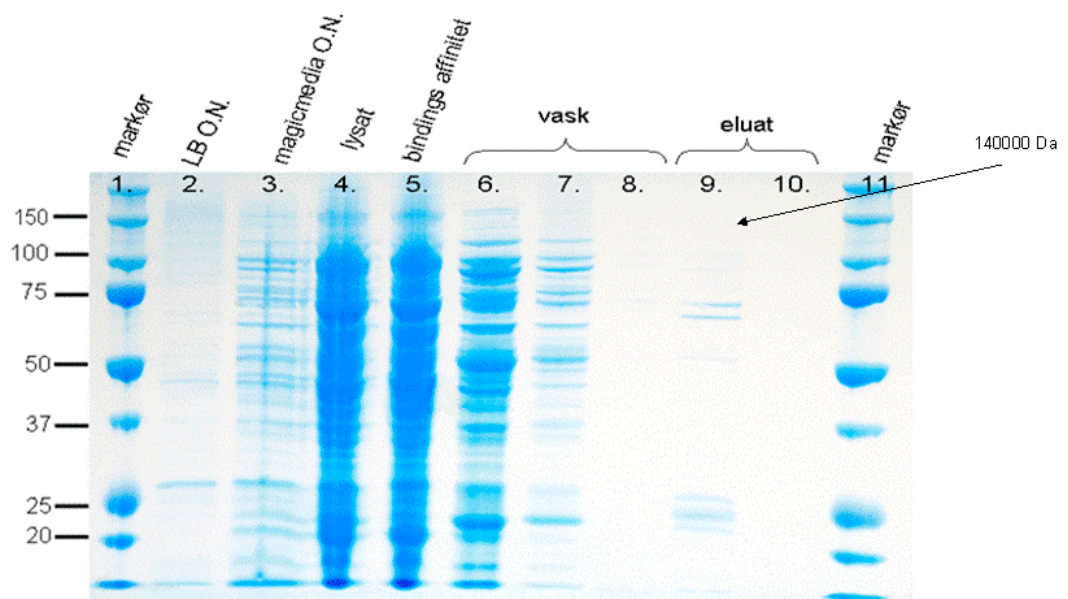
Det var ønskelig å gjenta proteinekspressjonen for Ahp, for å rense større mengder protein. Men dette var ikke like vellykket. Bilde under viser at det ikke er noe tydelig bånd i forventet størrelse (Figur 33).



**Figur 33. SDS-PAGE analyse av Ahp domene renset ved affinitetskromatografi. Forventet størrelse på proteinet 94000Da.**

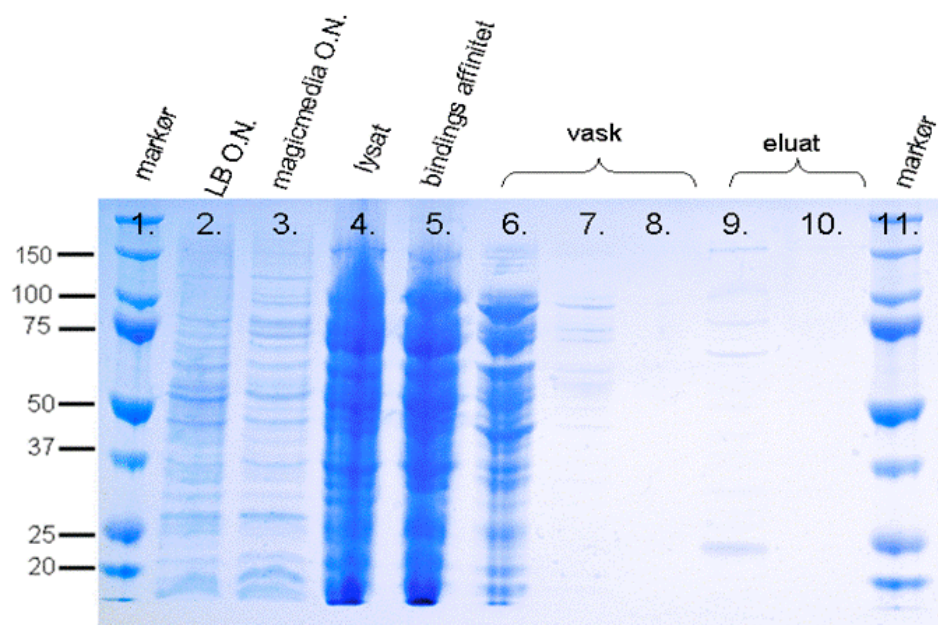
Som kontrollprotein var det ønskelig å bruke mcnE domene fra NC 172/5. mcnE er et større protein, noe som kan gjøre det vanskeligere å uttrykke og rense, så det ble gjort gjentatte forsøk på å rense det uten at det ga et tydelig bånd i ønsket størrelse (Figur 34).





**Figur 34. SDS\_PAGE analyse av mcnE domene renset med affinitetskromatografi, forventet størrelse på proteinet er 140.000Da.**

Celler transformert med pET21d vektoren uten innskudd ble også induisert med IPTG, og cellelysats renset på en nikkelkolonne (Figur 35).



**Figur 35. SDS-PAGE analyse av protein produsert i en celle transformert med pET21d vektor etter induksjon med IPTG.**

Eluatene ble lagret i 10 % glycerol ved - 20 °C, for at mulig aktivitet skulle bevares.

### 3.2.7 ATP/PPi utbytting

Målet med å uttrykke Ahp-domenet fra NC 172/5 var å finne ut hvilken aminosyre som domenet kan aktivere. Dette ble gjort ved å måle aminosyreavhengig ATP-pyrofosfat utbytting.

Reaksjonen ble utført ved å tilføre radioaktivt pyrofosfat og ATP til en løsning bestående av bla en aminosyre og det rensede proteinet. Dersom proteinet aktiverer aminosyren vil det dannes radioaktivt ATP. Reaksjonen blir tilført en stoppmiks som inneholder trekull og ikke-radioaktivt fosfat. Dersom proteinet aktiverer aminosyren og vi får dannet radioaktivt ATP, vil dette binde seg til trekullet, og vi kan måle aktiviteten med en scintillasjonsteller.

ATP-pyrofosfat utbyttings reaksjonen ble utført ved å teste alle tilgjengelige aminosyrer med proteinene, men ettersom det var små mengder rensset protein var det ikke mulig å gjenta forsøket med alle aminosyrene, så et utvalg av de mest sannsynlige ble foretatt. Resultatet er vist i tabellen under (tabell 8). Alle verdiene er stabilt lave, så ingen aktivitet er påvist.

**Tabell 8 ATP/PPi målinger med LSC**

	<b>Ahp (cpm)</b>	<b>mcnE (cpm)</b>	<b>pET21d (cpm)</b>
Glutamat	1089 1005 1389		1117 1119 1281
Glutamin	1174 1069 1137	1139 1054 952	1172 1180 1135
Ornitin	1149 995 1099		1226 1158 1226
Arginin	1215 1147 994		1032 976 1009
Isoleusin		966 1129 1097	1374 1453 1394
Alanin	1309 1200 1212	1175 1092 1141	1551 1845 1916
Blank	970	1082	1144

cpm= radioaktivitet målt per minutt  
tilsatt mengde radioaktivitet er 48000 cpm

## 4 Diskusjon

### 4.1 Identifisering av Ahp aktiverende domener hos cyanobakterier

Cyanobakteriestammer har i denne oppgaven blitt undersøkt i forsøk på å identifisere flere stammer som produserer cyanopeptolin. Cyanopeptolin syntetiseres ikke-ribosomalt og er karakteristisk fordi den inneholder den uvanlige aminosyren Ahp. Det har blitt undersøkt 52 cyanobakteriestammer, og det er bekreftet at 14 av disse har A-domener som aktiverer Ahp (Tabell 5). For å finne Ahp aktiverende domener er det brukt degenererte konsensusprimere som skiller mellom A-domener som aktiverer Ahp, og A-domener som aktiverer andre substrater (tabell 3). Det er hovedsakelig arter av *Microcystis*, *Anabaena* og noen *Planktothrix* som ga positive resultater i dette arbeidet, og ettersom primerene er laget ved å sammenligne kjente sekvenser av Ahp-aktiverende domener fra disse stammene, er det godt mulig at også andre av de undersøkte cyanobakteriestammer inneholder Ahp-domener som ikke fanges opp av primerene jeg har brukt. Det ble undersøkt flest stammer av *Microcystis* og *Anabaena* fordi de var tilgjengelige fra NIVA. Andre arter som det ble analysert var fra slektene *Nostoc*, *Phormidium*, *Merismopedia*, *Pseudoanabaena*, *Aphanisomenon* og *Pleurocapsa* (Tabell 5), noen av dem ga PCR-produkt i forventet størrelse, men sekvensene viste at de ikke stammet fra peptid syntetasegener.

De 14 sekvensene som er funnet her inneholder de konserverte områdene som er typiske for A-domener (Marahiel *et al.*, 1997). Figur 36 viser de konserverte områdene A3-A7 markert med gult. *In silico* analyse av det aktive bindingsetet viser at A-domenene har samme signatursekvens, noe som betyr at de antagelig aktiverer det samme substratet. Siden Ahp er karakteristisk for cyanopeptoliner (Namikoshi og Rinehart, 1996), er det trolig at vi her har identifisert nye stammer som syntetiserer cyanopeptoliner. Signatursekvensen for Ahp-aktiverende A-domener er DVENAGVVT (Tabell 6), og de angitte posisjonene er markert med røde stjerner i sekvenssammenstillingen (Figur 36). Det er mulig Ahp finnes i andre peptider som er

helt ulike cyanopeptolin, men det kan understrekes at jeg i denne oppgaven ikke har funnet noe Ahp som ikke hører til cyanopeptolin.

Det ble utført fylogenetiske analyser på de sekvensene som ble funnet her sammen med publiserte sekvenser fra A-domener hos *Anabaena* (Rouhiainen *et al.*, 2000), *Planktothrix* (Rounge *et al.*, 2007) og *Microcystis* (Tooming-Klunderud *et al.*, 2007). Figur 18 viser at A-domenesekvensene klustrer sammen avhengig hvilke aminosyrer de aktiverer, og Ahp danner en egen gruppe. Det fylogenetiske treet viser også at Ahp-sekvenser fra samme slekt har størst likhet, og det antydes at sekvensene fra *Microcystis* (mcn) og *Planktothrix* (oci) er nærmere beslektet hverandre enn de er med sekvensene fra *Anabaena*. Tilsvarende resultater er også vist av Rounge (2007).

Evolusjonært slektskap fremstilt som fylogenetiske trær viser ikke mulig rekombinasjon mellom de ulike stammene. Når ulike stammer rekombinerer sekvensene innbyrdes, vil hierarkiske trær ikke gi mening i og med at sekvensenes slektskapsmening er forbundet i et nettverk. Nettverket vist for *Microcystis* (Figur 19), og mosaikkmønsteret (Figur 22) viser klare tegn på rekombinasjon mellom artene. SplitsTrees nettverksanalyse bekrefter rekombinasjon for *Microcystis* som statistisk signifikant ved Phi-testen (Bruen *et al.*, 2006). De informative setene danner bokser av sekvenser, mosaikkstruktur, som viser at deler av sekvensen til NC 144 består av områder som kan stamme fra forskjellige forfedre. RDP2 analysen (Tabell 7) viser det samme som mosaikkstrukturen, og gir forslag til hvilke deler av sekvensen som kan stamme fra de andre *Microcystis*-stammene.

*Anabaena* 90 og NC 269/6 har svært homologe DNA-sekvensene, det er kun en base som er skiller de to. Denne ene basen kan komme av en mutasjon, og behøver ikke bety rekombinasjon. Men det er påfallende at de to stammer er så like. *Anabaena* 90 er fra Finland, mens NC 269/6 er fra Norge. Analyser av de informative setene (Figur 23) og SplitsTrees (Figur 20) viser at NC 419/2, NC 139 og NC 83/1 danner en egen gruppe. Det vises også i det fylogenetiske treet (Figur 18), at *Anabaena*-



stammene er delt i to grupper. Figuren fra SplitsTrees danner en liten antydning til nettverk. For NC 83/1 ser man på mosaikkmønsteret at det er en rekombinatorisk mulighet ved posisjon 154. Her har NC 83/1 samme base som NC 269/6 og *Anabaena* 90.

Hos *Planktothrix* (Figur 18, Figur 21, Figur 24) ser man at NC 116 og NIES 205 er nærmere beslektet enn NC 98. Det er vanskelig å si noe om mulig rekombinasjon utfra sekvensene til *Anabaena* og *Planktothrix*, og aalyser med RDP2 ga ingen resultat som viste rekombinasjon. Sekvenser fra flere stammer vil kunne gi ytterligere informasjon.

Cyanopeptolinsyntetaser kan være resultat av evolusjonære mekanismer som involverer horisontal genoverføring, ettersom strukturelt like peptider forekommer i forskjellige organismer. Horisontal genoverføring har blitt foreslått å være viktig evolusjonær mekanisme som er involvert i genomisk stabilitet og variasjon mellom nært beslektede cyanobakteriestammer (Rudi *et al.*, 1998).

## 4.2 Identifisering av forløperen til Ahp i cyanopeptolin

Reaksjonen som fører til dannelsen av Ahp, den uvanlige aminosyren i cyanopeptolin, er enda ikke kjent. Rouhiainen (2000) foreslår at glutamin kan være en mulig forløper siden Ahp-signatursekvensen (Tabell 6) ligner på signatursekvensen til A-domener som aktiverer glutamin. Ser man på signatursekvensene i Tabell 6, ser man at både, glutamin, glutamat og isoleucin har fem like aminosyrer i samme posisjon, ornitin har fire like, mens det for arginin ikke foreligger noen signatursekvens. Figur 7 viser at både glutamat, glutaminsyresemialdehyd, arginin eller ornitin er mulige forløpere for Ahp. Ahp grupperer seg ikke sammen med glutaminaktiverende domener i det fylogenetiske treet her, men danner en egen gruppe.

For å identifisere forløperen til Ahp, ble heterolog ekspresjon av proteinet utført. Flere forsøk på å rense proteinet under native forhold ble gjort. Første forsøk var

vellykket, men prøven ble ikke oppbevart forsvarlig for å bevare aktiviteten til enzymet. Dessverre viste det seg at det var svært vanskelig å gjenta ekspresjonen. (Magicmedia som ble benyttet har en holdbarhetsdato på 1 mnd, noe som viste seg å stemme), og nye forsøk med nytt medium ga heller ikke ønsket resultat. Ved ekspresjon av rekombinante proteiner kan det være at proteinet danner inklusjonslegemer, det vil si at mekanismen for proteinfolding forstyrres, og det gjør det vanskeligere å rense uten sterke detergenter som urea for å gjøre proteinet løselig. Ettersom det som regel vil være noen løselige proteiner tilstede, bør det være mulig å rense proteinet under native forhold. Siden det viste seg å være vanskelig å gjenta rensingen her, kan det være en mulighet å rense proteinet under denaturerende forhold. Det kan også forsøkes å variere temperaturen ved proteinekspresjon med lavere temperaturer, som ved 16-30 °C, i dette forsøket ble det brukt 37 °C.

Ettersom det var et tydelig bånd i forventet størrelse etter rensing av Ahp i det første forsøket (Figur 32), og noen antydninger til bånd i de andre forsøkene (Figur 33, Figur 34), ble enzymatisk aktivitet testet. Det ble ikke påvist noen aktivitet ved ATP-PPi utbyttingsreaksjonen (Tabell 8 ). Noe av grunnen til at det ikke ble påvist aktivitet til Ahp-proteinene som ble renses først kan være at enzymatisk aktivitet ikke ble bevart ved forsvarlig lagring. For de andre forsøkene ble prøvene lagret i 10 % glyserol, men her har nok ikke proteinet blitt tilstrekkelig uttrykt eller renses.

McnE-domenet klustrer med de andre isoleusin aktiverende A-domenene i det fylogenetiske treet (Figur 18), men hvilken aminosyre det aktiverer er fremdeles ukjent. Utfra signaturssekvensen (Tabell 6) ser det mer ut som det aktiverer glutamin (Tooming-Klunderud *et al.*, 2007). Sammenligner man signaturssekvensen med andre mulige substrater, er det 8 posisjoner som er like for glutamat og 7 for isoleusin. Disse er også mulige aminosyrer som aktiveres av A-domenet i mcnE. McnE viste seg å være vanskelig å uttrykke, forsøkene som ble gjort med dette proteinet for å identifisere hvilken aminosyre som aktiveres ga ingen resultat.

## 4.3 Videre forskning

Ikke-ribosomal peptidsyntetase katalyserer dannelsen av strukturelt forskjellige og biologisk viktige peptider. NRPS har et enormt potensial for fremtidig forskning til å generere nye bioaktive løsninger. Det vil for videre forskning være mulig med dagens teknologi å fortsette arbeidet med sekvensering av genomer. Jakten på sekundærmetabolitter og gener som koder for syntese av disse produktene ser ut til å være viktig arbeid i søken etter nye biologisk aktive substanser. Dette vil kunne gi verdifull kunnskap og forståelse av utbredelse og variasjoner mellom stammer av cyanobakterier, og hos andre fylum som sopp (Ehrenreich *et al.*, 2005). Med informasjonen som er tilgjengelig i dag er det mye som tyder på at det er like stor variasjon i de ikke-ribosomale peptidsyntetasene som produktene de lager. Sekvenserings teknologi er under stadig utvikling, og nye teknikker gir nye muligheter. 454 sekvensering ([www.454.com](http://www.454.com)) er en metode hvor man kan sekvensere hele 20 megabaser på 4,5 timer. Denne metoden gjør at man kan sekvensere hele genomer svært raskt. Det vil være interessant å finne flere cyanobakteriestammer som produserer cyanopeptolin, for å se nærmere på strukturen og funksjonen til peptidet, og gi en dypere innsikt i evolusjonære mekanismer. Kanskje finnes det andre peptider enn cyanopeptolin som har Ahp. Økte muligheter med sekvensering medfører også behov for heterolog ekspresjon av sekundærmetabolitter. En rekke strategier har blitt beskrevet, men det er hovedsakelig små gener som har blitt uttrykt. Ettersom forsøkene i denne oppgaven ikke ga forventede resultater vil det for fremtids forskning være interessant å undersøke dette område videre for å kunne si noe mer om hvordan syntesen av de ikke-ribosomale peptidene foregår. Massespektrometri blir brukt til å identifisere proteiner, og en slik metode kan også brukes til å identifisere forløperen til Ahp. Ved å bruke gassfase eliminasjonsreaksjonen som skjer under massespektrometri, kan man raskt karakterisere substrater, intermediater og produkter som er koblet til fosfopanteteinylarmen på PCP-domenet til NRPS systemet (Dorrestein *et al.*, 2006).

Bestemmelse av 3d-stukture av kjernedomener i kombinasjon med biokjemiske forsøk har gjort det mulig å forutsi substratspesifisiteten til A-domener, men det er

fremdeles ikke mulig å forutsi presis kjemisk struktur til produktene som er syntetisert av en ukjent NRPS. Det er ikke alltid mulig å kunne si noe om hvordan en modul eller et domene vil oppføre seg når det blir tatt ut av sin naturlige setting, det er behov for mer biokjemiske undersøkelser og strukturinformasjon som kan gi svar på hvordan nye sykliske peptider kan dannes. Hvilken rolle cyanopeptolin har på organismen som produserer den kan gjøres ved å forsøke å skru av produksjonen av cyanopeptolin, for så å kunne sammenligne om dette har noen effekt på organismen. Til nå har det vist seg å være vanskelig for NC 172/5 (Tooming-Klunderud *et al.*, 2007), men for stammer med *Microcystin* har det latt seg gjøre (Rohrlack *et al.*, 1999).

Den stadig økende mengde informasjon om cyanobakterienes peptider har gitt større innsikt i disse svært varierende produktene. Fremdeles er det mange uløste spørsmål. Det er kun noen få undersøkelser som er publisert om de ikke-ribosomale peptidenes økologiske rolle, og de fleste av disse undersøkelsene handler om microcystiner. Den store likheten i ulike NRPS gener og antydninger om rekombiasjonshendelser, kan tyde på at flere peptider har lignende funksjon, eller at de har utviklet seg i nær ko-evolusjon. Mer genetisk data vil gi en dypere innsikt i evolusjonen av genene.

## Referanser:

- Abascal, F., Zardoya, R., and Posada, D.** (2005). ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* **21**, 2104-2105.
- Adachi, J., and Hasegawa, M.** (1996). MOLPHY version 2.3: programs for molecular phylogenetics based on maximum likelihood. (Computer Science Monographs of Institute of Statistical Mathematics).
- Adachi, J., Waddell, P.J., Martin, W., and Hasegawa, M.** (2000). Plastid genome phylogeny and a model of amino acid substitution for proteins encoded by chloroplast DNA. *Journal of Molecular Evolution* **50**, 348-358.
- Akada.** (1994). Quick-check method to test the size of *Escherichia coli* plasmids. *BioTechniques* **17**, 58.
- Bister, B., Keller, S., Baumann, H.I., Nicholson, G., Weist, S., Jung, G., Süßmuth, R.D., and Jüttner, F.** (2004). Cyanopeptolin 963A, a Chymotrypsin Inhibitor of *Microcystis* PCC 7806. *Journal of Natural Products* **67**, 1755-1757.
- Boni, M.F., Posada, D., and Feldman, M.W.** (2007). An exact nonparametric method for inferring mosaic structure in sequence triplets. *Genetics*, Doi106.068874.
- Bruen, T.C., Philippe, H., and D., B.** (2006). A Simple and Robust Statistical Test for Detecting the Presence of Recombination. *Genetics* **172**, 2665-2681.
- Cao, Y., Janke, A., Waddell, P.J., Westerman, M., Tankenaka, O., Murata, S., Okada, N., Paabo, S., and Hasegawa, M.** (1998). Conflict among individual mitochondrial proteins in resolving the phylogeny of eutherian orders. *Journal of Molecular Evolution* **47**, 307-322.
- Carmichael, W.W., Azevedo, S.M.F.O., An, J.S., Molica, R.J.R., Jochimsen, E.M., Lau, S., Rinehart, K.L., Shaw, G.R., and K., E.G.** (2001). Human fatalities from cyanobacteria: chemical and biological evidence for cyanotoxins. *Environmental Health Perspectives* **109**, 663-668.
- Castenholz, R.W.** (2001). Phylum BX. Cyanobacteria. Oxygenic photosynthetic bacteria, . *Bergey's manual of systematic bacteriology* **1**, 473-599.
- Challis, G.L., Ravel, J., and Townsend, C.A.** (2000). Predictive, structure-based model of amino acid recognition by nonribosomal peptide synthetase adenylation domains. *Chemistry & Biology* **7**, 211-224.

- Chenna, R., Sugawara, H., Koike, T., Lopez, R., Gibson, T.J., Higgins, D.G., and Thompson, J.D.** (2003). Multiple sequence alignment with the Clustal series of programs. *Nucleic Acids Research* **31**, 3497-3500.
- Chomczynski, P., and Rymaszewski, M.** (2006). Alkaline polyethylene glycol-based method for direct PCR from bacteria, eukaryotic tissue samples, and whole blood. *BioTechniques* **40**, 454-458.
- Christiansen, G., Fastner, J., Erhard, M., Borner, T., and Dittmann, E.** (2003). Microcystin Biosynthesis in *Planktothrix*: Genes, Evolution, and Manipulation. *Journal of Bacteriology* **185**, 564-573.
- Dayhoff, M.O., Schwartz, R.M., and Orcutt, B.C.** (1978). A model of evolutionary change in proteins. *Atlas of Protein Sequence and Structure* **5**, 345-352.
- Dimmic, M.W., Rest, J.S., Mindell, D.P., and Goldstein, D.** (2002). rtREV: An amino acid substitution matrix for inference of retrovirus and reverse transcriptase phylogeny. *Journal of Molecular Evolution* **55**, 65-73.
- Dorrestein, P.C., Bumpus, S.B., Calderone, C.T., Garneau-Tsodikova, S., Aron, Z.D., Straight, P.D., Kolter, R., Walsh, C.T., and Kelleher, N.L.** (2006). Facile Detection of Acyl and Peptidyl Intermediates on Thiotemplate Carrier Domains via Phosphopantetheinyl Elimination Reactions during Tandem Mass Spectrometry. *American Chemical Society* **45**, 12756-12766.
- Döhren, H., Keller, U., Vater, J., and Zocher, R.** (1997). Multifunctional Peptide Synthetases. *American Chemical Society* **97**, 2675-2705.
- Ehrenreich, I.M., Waterbury, J.B., and A., W.E.** (2005). Distribution and Diversity of Natural Product Gene in Marine and Freshwater Cyanobacterial Cultures and Genomes. *Applied and Environmental Microbiology* **71**, 7401-7413.
- Felsenstein, J.** (1981). Evolutionary trees from DNA sequences - a maximum likelihood approach. *Journal of Molecular Evolution* **17**, 368-376.
- Felsenstein, J.** (1985). Confidence Limits on Phylogenies: An Approach Using the Bootstrap Evolution **39**, 783-791.
- Finking, R., and Marahiel, M.A.** (2004). Biosynthesis of nonribosomal peptides. *Annual Review of Microbiology* **58**, 453-488.
- Gold, H.S., and Moellering, R.C.** (1996). Antimicrobial-Drug Resistance. *The New England Journal of Medicine* **335**, 1445-1453.
- Guindon, S., and Gascuel, O.** (2003). PhyML - A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic Biology* **52**, 696-704.

- Harada, K.I., Mayumi, T., Shimada, T., Suzuki, M., Kondo, F., and Watanabe, M.F.** (1993). Occurrence of four depsipeptides, aeruginopeptins, together with microcystins from toxic cyanobacteria. *Tetrahedron Letters* **34**, 6091-6094.
- Hasegawa, M., Kishino, H., and Yano, T.** (1985). Dating the human-ape split by a molecular clock of mitochondrial DNA. *Journal of Molecular Evolution* **22**, 160-174.
- Heinkoff, S., and Heinkoff, J.G.** (1992). Amino acid substitution matrices from protein blocks. *Proceedings of the National Academy of Sciences* **89**, 10915-10919.
- Huelsenbeck, J.P., Ronquist, F., Nielsen, R., and P., B.J.** (2001). Bayesian Inference of Phylogeny and Its Impact on Evolutionary Biology. *Science* **294**, 2310-2314.
- Huson, D.H., and Bryant, D.** (2006). Application of Phylogenetic Networks in Evolutionary Studies. *Molecular Biology and Evolution* **23**, 254-267.
- Inoue.** (1990). *E. coli* competent cells and transformation. *Gene* **96**, 23-28.
- Jones, D.T., Taylor, W.R., and Thornton, J.M.** (1992). The rapid generation of mutation data matrices from protein sequences. *Computer Applications in the Biosciences* **8**, 275-282.
- Jukes, T., and Cantor, C.** (1969). *Evolution of protein molecules.* (Academic Press, New York).
- Kass, R.E., and Raftery, A.E.** (1995). Bayes factors. *Journal of the American Statistical Association* **90**, 773-795.
- Kimura, M.** (1980). A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution* **16**, 111-120.
- Lanave, C., Preparata, G., Saccone, C., and Serio, G.** (1984). A new method for calculating evolutionary substitution rates. *Journal of Molecular Evolution* **20**, 86-93.
- Lautru, S., and Challis, G.L.** (2004). Substrate recognition by nonribosomal peptide synthetase multi-enzymes. *Microbiology* **150**, 1629-1636.
- Lee, S.G., and Lipman, F.** (1975). Tyrocidine synthetase system. *Methods in Enzymology* **43**, 585-602.
- Lewis, P.O.** (2001). A Likelihood Approach to Estimating Phylogeny from Discrete Morphological Character Data. *Systematic Biology* **50**, 913-925.

- Linne, U., and Marahiel, M.A.** (2004). Reactions Catalyzed by Mature and Recombinant Nonribosomal Peptide Synthetases. *Methods in Enzymology* **388**, 293-315.
- Marahiel, M.A., Stachelhaus, T., and Mootz, H.D.** (1997). Modular Peptide Synthetases Involved in Nonribosomal Peptide Synthesis. *Chemical Reviews* **97**, 2651-2673.
- Martin, C., Oberer, L., Ino, T., Konig, W., Busch, M., and Weckesser, J.** (1993). Cyanopeptolins, new depsipeptides from the cyanobacterium *Microcystis* sp. PCC 7806. *The Journal of Antibiotics* **46**, 1550-1556.
- Martin, D.P., Williamson, C., and Posada, D.** (2005). RDP2: recombination detection and analysis from sequence alignments. *Bioinformatics* **21**, 260-262.
- Mikalsen, B., Boison, G., Skulberg, O.M., Fastner, J., Davies, W., Gabrielsen, T.M., Rudi, K., and Jakobsen, K.S.** (2003). Natural variation in the microcystin synthetase operon *mcyABC* and impact on microcystin production in *Microcystis* strains. *Journal of Bacteriology* **185**, 2774-2785.
- Moffitt, M.C., and Neilan, B.A.** (2004). Characterization of the nodularin synthetase gene cluster and proposed theory of the evolution of cyanobacterial hepatotoxins. *Applied and Environmental Microbiology* **70**, 6353-6362.
- Mootz, H.D., Schwarzer, D., and Marahiel, M.A.** (2002). Ways of Assembling Complex Natural Products on Modular Nonribosomal Peptide Synthetases. *ChemBioChem : a European Journal of Chemical Biology* **3**, 490-504.
- Muller, T., and Vingron, M.** (2000). Modeling aminoacid replacement. *Journal of Computational Biology* **7**, 761-776.
- Namikoshi, M., and Rinehart, K.L.** (1996). Bioactive compounds produced by cyanobacteria. *Journal of Industrial Microbiology* **17**, 373-384.
- Nishizawa, T., Ueda, A., Asayama, M., Fujii, K., Harada, K., Ochi, K., and Shirai, M.** (2000). Polyketide Synthase Gene Coupled to the Peptide Synthetase Module Involved in the Biosynthesis of the Cyclic Heptapeptide Microcystin. *The Japanese Biochemical Society* **127**.
- Nylander, J.A.A., Ronquist, F., Huelsenbeck, J.P., and Luis, J.** (2004). Bayesian Phylogenetic Analysis of Combined Data. *Society of Systematic Biologists* **53**, 47-67.
- Okino, T., Murakami, M., Haraguchi, R., Munekata, H., Matasuda, H., and Yamaguchi, K.** (1993). Micropeptins A and B, plasmin and trypsin inhibitors from the blue-green alga *Microcystis aeruginosa*. *Tetrahedron Letters* **34**, 8131-8134.



- Posada, D., and Crandall, K.A.** (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817-818.
- Rippka, R., Deruelles, J., Waterbury, J.B., Herdman, M., and Stanier, R.Y.** (1979). Generic Assignments, Strain Histories and Properties of Pure Cultures of Cyanobacteria. *Journal of General Microbiology* **111**, 1-61.
- Rohrlack, T., Dittmann, E., Henning, M., Börner, T., and Kohl, J.G.** (1999). Role of Microcystins in Poisoning and Food Ingestion Inhibition of *Daphnia galeata* Caused by Cyanobacterium *Microcystis aeruginosa*. *Applied and environmental microbiology* **65**, 737-739.
- Ronquist, F., and Huelsenbeck, J.P.** (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics Applications Note* **19**, 1572-1574.
- Rouhiainen, L., Vakkilainen, T., Lumbye Siemer, B.L., Buikema, W., Haselkorn, R., and Sivonen, K.** (2004). Genes Coding for Hepatotoxic Heptapeptides (Microcystins) in the Cyanobacterium *Anabaena Strain 90*. *Applied and environmental microbiology* **70**, 686-692.
- Rouhiainen, L., Paulin, L., Suomalainen, S., Hyytiäinen, H., Buikema, W., Haselkorn, R., and Sivonen, K.** (2000). Genes encoding synthetases of cyclic depsipeptides, anabaenopeptilides, in *Anabaena strain 90*. *Molecular Microbiology* **37**, 156-167.
- Rounge, T.B., Rohrlack, T., Tooming-Klunderud, A., Kristensen, A., and Jakobsen, K.S.** (2007). Comparison of cyanopeptolin genes in *Planktothrix*, *Microcystis* and *Anabaena*: evidence for independent evolution within each genus. submitted.
- Rudi, K., Skulberg, O.M., and Jakobsen, K.S.** (1998). Evolution of Cyanobacteria by Exchange of Genetic Material among Phyletically Related Strains. *Journal of Bacteriology* **180**, 3453-3461.
- Saitou, N., and Nei, M.** (1987). The Neighbor-Joining Method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406-425.
- Sambrook, J., and Russell, W.R.** (2001). *Molecular Cloning, a laboratory manual*. (Cold spring harbor laboratory press).
- Sanger, F., Nicklen, S., and R., C.A.** (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences* **74**, 5463-5467.

- Schwarzer, D., Finking, R., and Marahiel, M.A.** (2003). Nonribosomal peptides: from genes to products. *The Royal Society of Chemistry* **20**, 278-287.
- Sieber, S.A., and Marahiel, M.A.** (2003). Learning from nature's drug factories: Nonribosomal synthesis of macrocyclic peptides. *Journal of Bacteriology* **185**, 7036-7043.
- Stachelhaus, T., and Marahiel, M.A.** (1995). Modular Structure of peptide synthetases revealed by dissection of the multifunctional enzyme GrsA. *The Journal of Biological Chemistry* **270**.
- Stachelhaus, T., Mootz, H.D., and Marahiel, M.A.** (1999). The specificity-conferring code of adenylation domains in nonribosomal peptide synthetases. *Chemistry & Biology* **6**, 493-505.
- Stachelhaus, T., Mootz, H.D., Bergendahl, V., and Marahiel, M.A.** (1998). Peptide Bond Formation in Nonribosomal Peptide Biosynthesis. Catalytic role of the condensation domain. *Journal of Biological Chemistry* **273**, 22773-22781.
- Thompson, J.D., Higgins, D.G., and Gibson, T.J.** (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673-4680.
- Tooming-Klunderud, A., Rohrlack, T., Shalchian-Tabrizi, K., Kristensen, T., and Jakobsen, K.S.** (2007). Structural analysis of a non-ribosomal halogenated cyclic peptide and its putative operon from *Microcystis*: implications for evolution of cyanopeptolins. *Microbiology* **153**.
- Tsukamoto, S., Painuly, P., Young, K.A., Yang, X.M., Shimizu, Y., and Cornell, L.** (1993). Microcystilide-a - a Novel Cell-Differentiation-Promoting Dipeptide from *Microcystis aeruginosa* No-15-1840. *Journal of the American Chemical Society* **115**, 11046-11047.
- Vanderheiden, G.J., Fairchild, A.C., and R., J.G.** (1970). Construction of a Laboratory Press for Use with the French Pressure Cell. *Applied Microbiology* **19**, 875-877.
- Welker, M., and Döhren, H.** (2006). Cyanobacterial peptides - Nature's own combinatorial biosynthesis. *Federation of European Microbiological Societies* **30**, 530-563.
- Welker, M., Döhren, H., Täschner, H., Steinberg, C.E.W., and Erhard, M.** (2003). Toxic Microcystin in shallow lake Müggelsee (Germany) - dynamics, distribution, diversity. *Archives of Hydrobiology* **157**, 227-248.

- Whelan, S., and Goldman, N.** (2001). A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach. *Molecular Biology and Evolution* **18**, 691-699.
- Wilmotte, A., and Herdmann, M.** (2001). Phylogenetic relationships among cyanobacteria based on 16S rRNA sequences. *Bergey's manual of systematic bacteriology* **1**, 487-493.
- Yang, Z.** (2000). Maximum Likelihood Estimation on Large Phylogenies and Analysis of Adaptive Evolution in Human Influenza Virus A Springer New York **5**, 423-432.

## 5 Vedlegg

### 5.1 Sekvenssammmentilling av A-domener

Adp er sekvenser fra *Anabaena* 90 (Rouhiainen *et al.*, 2000), Mcn er sekvenser fra NC 172/5 *Microcystis cf. Wessenbergii*, (Tooming-Klunderud *et al.*, 2007) 205oci og 116oci er sekvenser fra *Planktothrix* (Rounge *et al.*, 2007), PCC 7806, NC 357, NC 171/10, NC 419, NC 269/6, NC 22, NC 139, NC 98, 228, NC 144, NC 83/1, NC 161/1 og NC 43 er funnet i denne oppgaven. Områder markert med gult viser konserverte områder (Marahiel *et al.*, 1997), mens røde stjerner viser antatte aminosyrebindingsete posisjoner (Stachelhaus *et al.*, 1999).

**Figur 36. Sekvenssammmentilling av A domener**

	..... .....	..... .....	..... .....	..... .....	..... .....
	10	20	30	40	50
<b>AdpA1_Gln</b>	TFMLADAQVS	VLLTQQHLVE	KLP----ENQ	EPVVCLDTDW	LVICESSQES
<b>AdpB1_Hty</b>	QFMLKDSGTS	VLLTQSFLLD	QLPIAEQQNP	YQVICLDQES	--FSSELTDN
<b>AdpB2_Ahp</b>	ADISEDQFS	VLITQQKLLN	SLP----VQG	VKIIVLDTES	EMLTNQSQVN
<b>AdpB3_Thr</b>	NLILEDQVS	VLLIQQHLVE	KLQ----QHQ	AHIVCLDSDG	EKIAQNSNSN
<b>AdpB4_Tyr</b>	SLMLEDQVS	LLLTQWTLIE	KLP----QHP	AKPVFLDEIW	QQIAQNSQEN
<b>AdpD_Ile</b>	RFMWADAQVS	VLLTQQHLVE	KLP----EHQ	VPVVCLDTDW	LVICESSPES
<b>McnA_Gln</b>	IFMLEDAAVK	VLLTQQKLIN	KLP----EHQ	AQLICLDADW	ELISQFSQDN
<b>McnB_Thr</b>	SFMLEDAQLS	VLLTQQKLGE	TLP----QHQ	AQIIYLDSDW	EKIAENSQSN
<b>AdpA2_Thr</b>	SFILTDTQVK	VLLTQQQLVN	KLP----AHT	AQLVCLDIDL	EKITQNSNSN
<b>205ociA2_T</b>	NFMLEDSQVK	VLVTQAKLVE	SIP----QHQ	AQLICLDTDW	EKIAQNITSN
<b>116ociA2_T</b>	NFMLEDSQVK	VLVTQAKLVE	SIP----QHQ	AQLICLDTDW	EKIAQNITSN
<b>McnC1_Ile</b>	LYILQDAQIR	VLLTQKHLIA	GLP----KHQ	SEIVCLDTDW	SFITQAQESN
<b>McnC2_Ahp</b>	TYMAEDAQIS	LLLTQESLLD	SLP----AED	VGIVVLDKLA	EILTVQSPEN
<b>McnC3_Phe</b>	QFAIADAQLS	LLLTQEGLID	KLP----EHQ	ARLILLDRFT	EEINQNSQDN
<b>McnC4_Tyr</b>	QFMLEDQVP	LLLTQSSLID	KLP----NHQ	AKTVFLAEIW	QEIKPMNRDN
<b>McnE_Gln</b>	SFMLQDTQVK	ILLTCESLQN	FLP----NHQ	AIVVCLDKDW	QQINQASLEN
<b>PCC7806_Ah</b>	TYMAEDAQIS	LLLTQESLLD	SLP----AED	VGIVVLDKLA	EKLTQVSREN
<b>NC357_Ahp</b>	TYMAEDAQIS	LLLTQESLLD	SLP----AED	VGIVVLDKLA	EILTVQSPEN
<b>NC171_10_A</b>	TYMAEDAQIS	LRLTQESLLD	SLP----VED	VGIIVLDKLA	EILTVQSPEN
<b>NC419_Ahp</b>	AGISEDQFS	VLITQQKLLN	SLP----VQG	VKVIVLDRES	EMLTNQSPAN
<b>NC269_6_Ah</b>	ADISEDQFS	VLITQQKLLN	SLP----VQG	VKIIVLDTES	EMLTNQSQVN
<b>NC22_Ahp</b>	TYMAEDAQIS	LLLTQESLLD	SLP----VED	VGIVVLDKLA	EKLTVESSEN
<b>NC139_Ahp</b>	AGISEDQFS	VLITQQKLLN	SLP----AQG	VKVIVLDRES	EMLTNQSPAN
<b>205ociA3_H</b>	TYLLEDTQLN	ILLTQSQFLG	QLP----QFQ	GTTICLDQDW	SIIANQSTVN
<b>205ociA5_I</b>	SFMLQDTQVK	IILLTCESLQN	SLQ----NHQ	TIVICLDKDW	QQINQASLEN
<b>205ociA6_P</b>	QFMLEDQIP	LLLTQHSLID	KLP----NHQ	GQTLFLEEIW	EKTAQPSQDN
<b>205ociA7_I</b>	SFMLQDTQVK	ILLTQQSLLP	SLP----NHQ	AIIVCLDKDW	KQINQ---EN
<b>116ociA1_H</b>	QFIIEDSGIL	LLLTQGFLKE	KLLLTNLTNL	NQIILYLDQES	--FSLLLTND
<b>116ociA3_H</b>	TYLLEDTQLN	ILLTQSQFLG	QLP----QFQ	GTTICLDQDW	SIIANQSTVN

116oci_A4_	SYMAEDAKIS	LLLTQQSLLS	SLP----VD-	VGIVVLDKSA	ESLTVQSLEN
116ociA5_I	SFMLQDTQVK	IILTCESLQN	SLQ----NHQ	TIVICLDKDW	QQINQASLEN
116ociA6_P	QFMLEDSQIP	LLLTQHSLID	KLP----NHQ	GQTLFLEEIW	EKTAQPSQDN
116ociA7_I	SFMLQDTQVK	IILTQQSLLP	SLP----NHQ	AIIVCLDKDW	KQINQ---EN
NC98_Ahp	TSMAEDAKIS	LLLTQQSLLS	SLP----VED	VGIVVLDKSA	EILTVQSMEN
NC228_1_Ah	TYMAEDAQIS	LLLTQESLLD	SLP----VED	VGIVVLDKLA	EKLTVESSEN
NC144_Ahp	TYMAEDAQIS	LLLTQESLLD	SLP----VED	VGIVVLDKLA	EKLTVGSSSEN
NC83_1_Ahp	ADISEDSQFS	VLITQQKLLN	SLP----VQG	VKVIVLDRES	EMLTNQSPAN
NC161_1_Ah	TYMAEDAQIS	LLLTQESLLD	SLP----VED	VGIVVLDKLA	EKLTVESSEN
NC43_Ahp	TYMAEDAQIS	LLLTQESLLD	SLP----VED	VGIVVLDKLA	EKLTVESSEN
205ociA4_A	SYMAEDAKIS	LLLTQQSLLS	SLP----VED	VGIVVLDKSA	ESLTVQSLEN
205ociA1_H	QFIIEDSGIL	LLLTQGFLKE	KLLLTNLTNL	NQILYLDEQS	--FSLLLTDN

....|....| ....|....| ....|....| ....|....| ....|....|  
60 70 80 90 100

## A3 (core 2)

AdpA1_Gln	PITEVQPGNL	AYVIYTSGST	GTPKGVMLSH	SNLCNHM---	SWMQATFPLT
AdpB1_Hty	PNPQSTPNNL	AYVIYTSGST	GRPKGVMLSH	QALVNLS---	LAWCKTFQVQ
AdpB2_Ahp	PVSEVKPENL	ACLLYTSGST	GKPKGVMLTH	AALVNHS---	SAISEVFGLT
AdpB3_Thr	PLNIATPSNL	AYVIYTSGST	GKPKGVMLNH	SHVVRLE---	AATDSWYNFN
AdpB4_Tyr	LTGIVTAANL	ANVIYTSGST	GKPKGVMLNH	QGLCNLA---	QAQIQAFGLH
AdpD_Ile	PITEVQPGNL	AYVIYTSGST	GTPKGVVNVH	QAVNRLV---	-KNTNYVQLT
McnA_Gln	PITDVQATNL	AYVIYTSGST	GQPKGVMLSH	SNLSNHM---	FWMQETFPLT
McnB_Thr	PENTVTPDNL	AYVIYTSGST	GKPKGVMLNH	SNVVRLE---	AATDAWYNFN
AdpA2_Thr	PVNTATSPNL	AYVIYTSGST	GQPKGVMLNH	HHVTRLE---	AATNSWYKFN
205ociA2_T	PESGVKPDNL	TYIIYTSGST	GKPKGVMLNH	ANVVRLE---	AATDSWYHFN
116ociA2_T	PESGVKPDNL	TYIIYTSGST	GKPKGVMLNH	ANVVRLE---	AATDSWYHFN
McnC1_Ile	PCCEVQPENL	AYAIYTSGST	GKPKGVMIQH	SSVVNLAHGL	HEAIYAHHQI
McnC2_Ahp	PLSEVAPENL	LCLLYTSGST	GKPKGVMLTH	AALVNHS---	WGISEVFGLT
McnC3_Phe	LTGVVTANNL	ANLIYTSGST	GKPKGVMLNH	RGLFNLA---	QAQIKTFNLQ
McnC4_Tyr	LTGKVTASNL	ANVIYTSGST	GKPKGVMLNH	QGLSNLA---	LAQIQTFAVS
McnE_Gln	LNSTVSADNL	AYVIYTSGST	GIPKGVIVTH	QAVNRLV---	-LNTNYIQFT
PCC7806_Ah	PLSEVVPENL	LCLLYTSGST	GKPKGVMLTH	AALVNHS---	WGISEVGLT
NC357_Ahp	PFSEVAPENL	LCLLYTSGST	GKPKGVMLTH	AALVNHS---	CGISKVFGLT
NC171_10_A	PFSEVAPENL	LCVLYTSGST	GKPKGVMLTH	AALVNHS---	WGISEVFGLT
NC419_Ahp	PVSEVKPENL	TCLLYTSGST	GKPKGVMLTH	AALVNHS---	SAISEVFALT
NC269_6_Ah	PVSEVKPENL	ACLLYTSGST	GKPKGVMLTH	AALVNHS---	SAISEVFGLT
NC22_Ahp	PLSEVAPENL	LCPLYTSGST	GKPKGVMLTH	AALVNHS---	WGISEVFGLT
NC139_Ahp	PVSEVKPENL	TCLLYTSGST	GKPKGVMLTH	AALVNHS---	SAISEVFALT
205ociA3_H	PIVEVDQHNL	AYIIYTSGST	GKPKGVMLNH	RSIVNYI---	LTAIRECNFT
205ociA5_I	LNSTVSADNL	AYVIYTSGST	GIPKGVIVTH	QAVNRLV---	-LNTNYIQFT
205ociA6_P	LTAKVTPSNL	ANVIYTSGST	GKPKGVMLNH	KGLCNLA---	QAQIETLGVH
205ociA7_I	LNSAVSADNL	AYIIYTSGST	GTPKGVVVTH	QAVNRLI---	-LNTNYIQFT
116ociA1_H	PSSQTKTDSL	AHVIYTSGST	GKPKGVMIEN	RSIVNLA---	LAWSDVFIQI
116ociA3_H	PIVEVDQHNL	AYIIYTSGST	GKPKGVMLNH	RSIVNYI---	LTAIRECNFT
116oci_A4_	PVSEVVPENL	LCVLYTSGST	GKPKGVMLTH	EALVNHS---	WGISEIFGLT
116ociA5_I	LNSTVSADNL	AYVIYTSGST	GIPKGVIVTH	QAVNRLV---	-LNTNYIQFT
116ociA6_P	LTAKVTPSNL	ANVIYTSGST	GKPKGVMLNH	KGLCNLA---	QAQIETLGVH
116ociA7_I	LNSAVSADNL	AYIIYTSGST	GTPKGVVVTH	QAVNRLI---	-LNTNYIQFT
NC98_Ahp	PVSEVIPENL	LCVLYTSGST	GKPKGVMLTH	EALVNHS---	WGISEIFGLT
NC228_1_Ah	PLSEVAPENL	LCLLYTSGST	GKPKGVMLTH	AALVNHS---	WGISEVFGLT
NC144_Ahp	PLSEVAPENL	LCLLYTSGST	GKPKGVMLTH	AALVNHS---	WGISEVFGLT
NC83_1_Ahp	PVSEVKPENL	TCLLYTSGST	GKPKGVMLTH	AALVNHS---	SAISEVFALT
NC161_1_Ah	PLSEVAPENL	LCLLYTSGST	GKPKGVMLTH	AALVNHS---	WGISEVFGLT
NC43_Ahp	PLSEVAPENL	LCLLYTSGST	GKPKGVMLTH	AALVNHS---	WGISEVFGLT
205ociA4_A	PVSEVVPENL	LCVLYTSGST	GKPKGVMLTH	EALVNHS---	WGISEIFGLT
205ociA1_H	PSSQTKTDSL	AYVIYTSGST	GKPKGVMIEN	RSIVNLA---	LAWSDVFIQI

....|....| ....|....| ....|....| ....|....| ....|....|  
110 120 130 140 150

## A4

AdpA1_Gln	EKDKVLQKTP	FGFDASVWEF	YAPLLAGGQL	LIAKPGGHTD	SAYLLRLIAQ
AdpB1_Hty	SQSRLQFGS	FSFDLSVAEI	ATTYVSGACL	YLPNIETLLP	SQVLVDFLAD
AdpB2_Ahp	SGDRVLQFAS	FSFDVAAEEI	FPTWYKGATV	VLRPTQMFPD	FASLSQFIAQ
AdpB3_Thr	SQDVWTFHFS	YAFDFSVWEV	WGALLYGGRL	VVVGVLVTRS	PKSFYELLCQ
AdpB4_Tyr	SDSRVLQFAS	LSFDACISEI	LMSLGSGGTL	YLGSKDSLMP	GTPLIERLRD
AdpD_Ile	PDDRVAQAAN	IAFDAATFEI	WGALLNGAKL	VMITKSVLLS	PQEFAANIRD
McnA_Gln	KTDRVLQKTP	FSFDASVWEF	YAPLLVGGQL	LIAQPGGHTD	SDYLLKTIAQ
McnB_Thr	SQDVWSLFHS	YAFDFSVWEM	WGALLYGGRL	VVVPYLITRS	PEAFYQLLCQ
AdpA2_Thr	SQDVWTFHFS	YAFDFSVWEI	WGALLYGGRL	VVVPYLLTRS	PESFYKLLSQ
205ociA2_T	SQDVWTLFHS	YAFDFSVWEM	WGALLYGGRL	VIVPYLVTRS	PESFYELLCQ
116ociA2_T	SQDVWTLFHS	YAFDFSVWEM	WGALLYGGRL	VIVPYLVTRS	PESFYELLCQ
McnC1_Ile	SPLRV SINAS	LAFDA SVQOI	IQ-LIYGHS	NIVPEKVRFD	SHAMLSHLRQ
McnC2_Ahp	ESDRVLQFAS	FGFDVAAEEI	FPTWLKGGTV	VLRPGQMFPD	LTDFA DFIEQ
McnC3_Phe	SNSRVLQFAS	LNFDASIWEI	IMAIGAGATL	YLDTKDALMP	GLPLMERLRD
McnC4_Tyr	SQSRVLQFAS	FSFDACISEV	LMALGAGATL	YLASKENLMP	GLPLIKQLRE
McnE_Gln	PDDRVLQASN	IAFDAATFEI	WGALLNGAKI	IIIAKSVLLS	PQELALSKE
PCC7806_Ah	ESDRVLQFAS	FGFDVAAEEI	FPTWLKGGTV	VLRPGQMFPD	LTDFA DFIEQ
NC357_Ahp	ESDRVLQFAS	FGFDVAAEEI	FPTWLKGGTV	VLRPGQMFPD	LTDFA DFIEQ
NC171_10_A	ESDRVLQFAS	FGFDVAAEEI	FPTWLKGGTV	VLRPGQMFPD	LTDFA DFIEQ
NC419_Ahp	SADRVLQFAS	FSFDVAAEEI	FPTWYKGATV	VLRPTQMFPD	FASLSQFIAQ
NC269_6_Ah	SGDRVLQFAS	FSFDVAAEEI	FPTWYKGATV	VLRPTQMFPD	FASLSQFIAQ
NC22_Ahp	ESDRVLQFAS	FGFDVAAEEI	FPTWLKGGTV	VLRPGQVFPT	LTDFA DFIEQ
NC139_Ahp	SADRVLQFAS	FSFDVAAEEI	FPTWYKGATV	VLRPTQMFPD	FASLSQFIAQ
205ociA3_H	SDDQILQFSS	ICFDASVEEI	FCGLSSGATL	VLRTEKMLRS	SEDFWQCCQK
205ociA5_I	PDDRVLQASN	IAFDAATFEI	WGALLNGAKI	IIIAKSVLLS	PQELALSKE
205ociA6_P	PSSRVLQFAS	FSFDACIWEI	LMALGAGATL	YLGTKREALMP	GLPLMERLRD
205ociA7_I	PDDCVVQASN	IAFDAATFEI	WGALLNGAKI	IIILPKSVLLS	PQEFALSKE
116ociA1_H	NNSRVLQFGS	LSFDLSIGE	ATALAAGASL	YLAQKETLLP	SQILVDFLTE
116ociA3_H	SDDQILQFSS	ICFDASVQEI	FSCLLSGAIL	VLRTEKMLRS	SEDFWQYCCQK
116ociA4	ESDRVLQFAS	FGFDVAAEEI	FPTWLKGGTV	VLRPWQMFPD	LTDFA DFIEQ
116ociA5_I	PDDRVLQASN	IAFDAATFEI	WGALLNGAKI	IIIAKSVLLS	PQELALSKE
116ociA6_P	PSSRVLQFAS	FSFDACIWEI	LMALGAGATL	YLGTKREALMP	GLPLMERLRD
116ociA7_I	PDDCVVQASN	IAFDAATFEI	WGALLNGAKI	IIILPKSVLLS	PQEFALSKE
NC98_Ahp	ESDRVLQFAS	FGFDVAAEEI	FPTWLKGGTV	VLRPWQMFPD	LTDFA DFIEQ
NC228_1_Ah	ESDRVLQFAS	FGFDVAAEEI	FPTWLKGGTV	VLRPGQMFPD	LTDFA DFIEQ
NC144_Ahp	ESDRVLQFAS	FGFDVAAEEI	FPTWLKGGTV	VLRPGQMFPD	LTDFA DFIEQ
NC83_1_Ahp	SADRVLQFAS	FSFDVAAEEI	FPTWYKGATV	VLRPTQMFPD	FASLSQFIAQ
NC161_1_Ah	ESDRVLQFAS	FGFDVAAEEI	FPTWLKGGTV	VLRPGQMFPD	LTDFA DFIEQ
NC43_Ahp	ESDRVLQFAS	FGFDVAAEEI	FPTWLKGGTV	VLRPGQMFPD	LTDFA DFIEQ
205ociA4_A	ESDRVLQFAS	FGFDVAAEEI	FPTWLKGGTV	VLRPWQMFPD	LTDFA DFIEQ
205ociA1_H	NNSRVLQFGS	LSFDLSIGE	ATALAAGASL	YLAQKETLLP	SQILVDFLTE

\*\* \*

	.... ....	.... ....	.... ....	.... ....	.... ....
	160	170	180	190	200
AdpA1_Gln	QQVTIVQLVP	S---LLQMLL	EQG-----	-----GI	ETCHSLKHVF
AdpB1_Hty	HKISHSFLSP	S-----AL	SVL-----	-----PQ	ATLPDLQYLS
AdpB2_Ahp	ESLSVLNITP	AYWHEWAVAV	SQD-----	-----NA	TVPPSLRIVA
AdpB3_Thr	EKVITILNQTP	SAFRQLIPAE	QSI-----	-----AT	VGDLNLRIVI
AdpB4_Tyr	YAITHITLPP	S-----AL	AVL-----	-----HK	VELTTLQTII
AdpD_Ile	REVSVLFLTT	ALFN----QL	ASF-----	-----VP	QAFSSLRVLL
McnA_Gln	QQVTTVQLVP	S---LLQMLL	EQG-----	-----GI	ENCQLLKRVF
McnB_Thr	EKVITILNQTP	TAFRQLIQAE	ESLKGSFP--	LSRGDRSSTT	DNDLSLRIVI
AdpA2_Thr	EQVTILNQTP	SAFRQLIQAE	QSI-----	-----G	MSNLNLRIVI
205ociA2_T	EKVITILNQTP	TAFRQLIQAE	ESVKGDISPL	LREVGKDSVT	ETDLSLRIVI
116ociA2_T	EKVITILNQTP	TAFRQLIQAE	ESVKGDISPL	LREVGKDSVT	ETDLSLRIVI
McnC1_Ile	QKIDVLDCTP	S---QLRLLI	SVG-----	-----LL	ESDNHPQSVL
McnC2_Ahp	ESLTVLNITP	AYWHEWALAV	SQS-----	-----LA	TVPSSLRVVA
McnC3_Phe	YGITHVTLPP	S-----AL	AAL-----	-----PL	EEIPELQTII

A horizontal number line with major tick marks labeled 210, 220, 230, 240, and 250. Between each major tick mark, there are four smaller, unlabeled tick marks, dividing each 10-unit interval into five 2-unit intervals. Two red stars are positioned above the number line, each centered over a tick mark. The first star is located at the third small tick mark after 210, which represents the number 215. The second star is located at the third small tick mark before 250, which represents the number 245.

AdpA1_Gln	CGGEVLPVAL	LEG-----LL	SKLDVNLHNL	YGPTETCIDA	TFCNCQR--E
AdpB1_Hty	VGGEASSAEL	VE-----R	WGTGRRFFNA	YGPTESTVAA	TIAICQ---P
AdpB2_Ahp	VGGDAVLLET	VT---IWRQV	VGDRVNCLNV	YGPT-EASVTA	IVHDLLHPQS
AdpB3_Thr	FGGETLELNS	LQP--WFDRH	GDQSPQLVNM	YGITETTVHV	TYRPLS-KAD
AdpB4_Tyr	VAGEACAVEL	I-----KQ	WSAGRNFNA	YGPT-EGSVCA	TLAKCS---A
AdpD_Ile	FGGEAVDPQW	VQ----EVLE	KGAPKQLLHV	YGPTENTTFS	SWYLVEE-LT
McnA_Gln	CGGEILPVAL	QEK-----LL	SQLNVNLCNL	YGPT-ECCIDV	TFWNCQR--E
McnB_Thr	FGGEALEINS	LQP--WFQRH	GDQCPQLVNM	YGITETTVHV	TYRPLS-MTD
AdpA2_Thr	FGGEALELES	LQP--WFERH	GDQKPQLVNM	YGITETTVHV	TYRPLS-LDD
205ociA2_T	FGGESLEINS	LQP--WFDRH	GDQLPQLVNM	YGITETTVHV	TYRPLS-MAD
116ociA2_T	FGGESLEINS	LQP--WFDRH	GDQLPQLVNM	YGITETTVHV	TYRPLS-MAD
McnC1_Ile	VGGEAIDES-	----MWATLA	KAENINFYNV	YGPT-ECTVDS	TICLIT---A
McnC2_Ahp	VGGDAVLPET	VN---IWRQM	VGKRVQCINV	YGPT-EASVTA	IVHDLDDYQS
McnC3_Phe	VAGEACSEE-	-----LIKQ	WATGRNFFNA	YGPT-ESTVCA	TVEKWT---D
McnC4_Tyr	VAGEACSP-	-----LVKK	WSQGRNFFNG	YGPT-EGSVCA	TVAKCT---P
McnE_Gln	FGGEAVEPKW	VQ----EVLE	KGAPQRLHVV	YGPTENTTFS	SWYLVEN-VA
PCC7806_Ah	VGGDAVLPET	VN---IWRQM	VGKRVQCINV	YGPT-EASVTA	IVHDLDDYQS
NC357_Ahp	VGGDAVLPET	VN---IWRHM	VGKRVQCINV	YGPTGASVTA	IVHDLDDYQS
NC171_10_A	VGGDAVLPET	VN---IWRQM	VGKRVQCINV	YGPT-EASVTA	IVHDLDDYQS
NC419_Ahp	VGGDAVLLET	VT---IWRQV	VGDRVNCLNV	YGPT-EASVTA	IVHDLLHPQS
NC269_6_Ah	VGGDAVLLET	VT---IWRQV	VGDRVNCLNV	YGPT-EASVTA	IVHDLLHPQS
NC22_Ahp	VGGDAVLPET	VN---IWRQM	VGKRVQCINV	YGPT-EASVTA	IVHDLDDYQS
NC139_Ahp	VGGDAVLLET	VT---IWRQV	VGDRVNCLNV	YGPT-EASVTA	IVHDLLHPQS
205ociA3_H	VGGEAIQQSK	VKQWHQNTVN	YSPLPQLFNA	YGPT-EATIIT	TLCKFT---S
205ociA5_I	FGGEAVEPKW	VQ----EVLE	KGAPQRLHVV	YGPTENTTFS	SWYLVEN-VA
205ociA6_P	VAGEACSP-	-----LVKK	WSQGRNFFNG	YGPT-EGSVCT	AIAKCT---P
205ociA7_I	FGGEAVDPKW	VQ----EVRS	KGAPQRLHVV	YGPTENTTFS	SWYLVEN-LP
116ociA1_H	VGGEACTAK-	-----LITE	WSKGRNFFNA	YGPTETTVCA	TVARFT---K
116ociA3_H	VGGEAIOPSK	VKOWHONTVN	YSPLPOLFNA	YGPT-EATIIT	TLCKFT---S



A horizontal number line with major tick marks and labels at 260, 270, 280, and 300. There are minor tick marks every 10 units. Two red asterisks are placed above the line: one at 278 and one at 288.

## VPGELHISGA

.....|.....|.....|.....|.....|.....|.....|.....|.....|.....|

          310               320               330               340



**A7 (core4)**

AdpA1_Gln	ELTQEKFIAN	PFSTY-----	-----PG	SRLYK <b>Y</b> KTGDLA	RYLPNG
AdpB1_Hty	ELTQEKFIPN	PFDNS-KFKI	PYRNAKDVQN	SKLY <b>Y</b> KTGDLA	RYLPDG
AdpB2_Ahp	ELTEEFKIPN	PFSGSRGAEI	LANSQSPVPS	SRLYK <b>Y</b> KTGDLA	RYLPDG
AdpB3_Thr	ELTAQRFISN	PFNGNSE---	-----	QLLY <b>Y</b> KSGDLA	RYLPNG
AdpB4_Tyr	ELTQEKFISN	PFG-----	-----P	SKLY <b>Y</b> KTGDLG	RYLADG
AdpD_Ile	ELTQEKFIPN	PFD-----	-----N	SKLY <b>Y</b> KTGDLA	RYLPDG
McnA_Gln	ELTQEKFIPN	PFSNY-----	-----PD	SRLYK <b>Y</b> KTGDLA	RYLPDG
McnB_Thr	ELTKERFISS	PFEKDEVIPP	TPLNKGGNEP	SKLY <b>Y</b> KTGDLA	RYLPKG
AdpA2_Thr	ELTGQRFISN	PFQRSKGG--	-----	ERLY <b>Y</b> RTGDLA	RYLPNG
205ociA2_T	ELTTERFISS	PFEHS-----	-----	NKLY <b>Y</b> KTGDLA	RYWPNG
116ociA2_T	ELTTERFISS	PFEHS-----	-----	NKLY <b>Y</b> KTGDLA	RYWPNG
McnC1_Ile	ELTQEKFIPN	PFEKDEVIPP	TPLNKGGNEP	SKLY <b>Y</b> KTGDL <b>C</b>	RYLPDG
McnC2_Ahp	ELTAEKFIDN	PFAHS-----	-----PF	NRL <b>Y</b> HK <b>Y</b> TGDLA	RYLPDG
McnC3_Phe	ELTQEKFIPN	PFEKDEVIPP	TPLNKGGNEP	SKLY <b>Y</b> KTGDLG	RYLPDG
McnC4_Tyr	ELTAEKFISN	PFVS-LDPPL	TPLDKGGEQP	SKLY <b>Y</b> KTGD <b>E</b> A	RYLPDG
McnE_Gln	ELTQEKFIPN	PFEKDEVIPP	TPLNKGGNEP	SKLY <b>Y</b> KTEDLA	RYLPDG
PCC7806_Ah	ELTAEKFIDN	PFANA-----	-----AF	NRLY <b>Y</b> KTGDLA	RYLPDG
NC357_Ahp	ELTAEKFIDN	PFAHS-----	-----PF	NRLY <b>Y</b> KTGDLA	RYLPDG
NC171_10_A	ELTAEKFIDN	PFANA-----	-----AF	NRLY <b>Y</b> KTGDLA	RYLPDG
NC419_Ahp	ELTEEFKIPN	PFSGSRGAEI	LANSQSPVPS	SRLY <b>Y</b> KTGDLA	RYLPDG
NC269_6_Ah	ELTEEFKIPN	PFSGSRGAEI	LANSQSPVPS	SRLY <b>Y</b> KTGDLA	RYLPDG
NC22_Ahp	ELTAEKFIHN	PFTNA-----	-----AF	NRLY <b>Y</b> KTGDLA	RYLPDG
NC139_Ahp	ELTEEFKIPN	PFSGSRGAEI	LANSQSPVPS	SRLY <b>Y</b> KTGDLA	RYLPDG
205ociA3_H	ELTQEKFIPN	PFEEEGLI <b>P</b> P	NPINKGSNES	SKLY <b>Y</b> KTGDL <b>C</b>	RYLPDG
205ociA5_I	ELTQEKFIPN	PFEKDEVIPP	TPLNKDGNEP	SKLY <b>Y</b> KTGDLA	RYLPDG
205ociA6_P	ELTAEKFISN	PFVS-LDPPL	TPLDKGGEQP	SKLY <b>Y</b> KTGDLA	RYLPDG
205ociA7_I	ELTQEKFIPN	PFS-----	-----NPD	SRLY <b>Y</b> KTGDLA	RYLPDG
116ociA1_H	ELTQEKFIPN	PFSN-----	-----YPD	SRLY <b>Y</b> KTGDLA	RYLPDG
116ociA3_H	ELTQEKFIPN	PFDEEGLI <b>P</b> P	NPINKGSNES	SKLY <b>Y</b> KTGDL <b>C</b>	RYLPDG
116oci_A4_	ELTAEKFIDN	PFAAA-----	-----PF	NRLY <b>Y</b> KTGDLA	RYLPNG
116ociA5_I	ELTQEKFIPN	PFEKDEVIPP	TPLNKDGNEP	SKLY <b>Y</b> KTGDLA	RYLPDG
116ociA6_P	ELTAEKFISN	PFVS-LDPPL	TPLDKGGEQP	SKLY <b>Y</b> KTGDLA	RYLPDG
116ociA7_I	ELTQEKFIPN	PFS-----	-----NPD	SRLY <b>Y</b> KTGDLA	RYLPDG
NC98_Ahp	ELTAEKFIDN	PFADA-----	-----PF	NRLY <b>Y</b> KTGDLA	RYLPNG
NC228_1_Ah	ELTAEKFIHN	PFANA-----	-----AF	NRLY <b>Y</b> KTGDLA	RYLPDG
NC144_Ahp	ELTAEKFINN	PFAHS-----	-----PF	NRLY <b>Y</b> KTGDLA	RYLPDG
NC83_1_Ahp	ELTEEFKIPN	PFSGSRGAEI	LANSQSPVPS	SRLY <b>Y</b> KTGDLA	RYLPDG
NC161_1_Ah	ELTAEKFIHN	PFANA-----	-----AF	NRLY <b>Y</b> KTGDLA	RYLPDG
NC43_Ahp	ELTAEKFIHN	PFANA-----	-----AF	NRLY <b>Y</b> KTGDLA	RYLPDG
205ociA4_A	ELTAEKFIDN	PFAAA-----	-----PF	NRLY <b>Y</b> KTGDLA	RYLPNG
205ociA1_H	ELTQEKFIPN	LFDK-----	-----AEG	SKLY <b>Y</b> KTGDLA	RYLPDG

## 5.2 Sekvenseringsprimere

**Tabell 9 Sekvenseringsprimere:**

pUC/M13	5' - CGC CAG GGT TTT CCC AGT CAC GAC - 3'	Sekvenseringsprimer for pGEM vektor
SP6	5' - CAT ACG ATT TAG GTG ACA CTA TAG - 3'	Sekvenseringsprimer for pGEM vektor
T7	5' - TAA TAC GAC TCA CTA TAG GG - 3'	Sekvenseringsprimer for pET vektor
T7term	5' - GCT AGT TAT TGC TCA GCG G - 3'	Sekvenseringsprimer for pET vektor
mcnEintF1	5' - CAA GGT TCA TCA AAG AAT GA - 3'	sekvensering av mcnE
mcnEintF2	5' - GGC AAT TGA GAA ATT AAT GGG CTT T - 3'	sekvensering av mcnE
mcnEintF3	5' - GTG GTA TGT TCA GAG CAA AA - 3'	sekvensering av mcnE
mcnEint4F	5' - GGC ACTA AG CCT AAA GGA AAA T - 3'	sekvensering av mcnE
mcnEint5F	5' - GAG GCT GTA CTT AGT CAA AA - 3'	sekvensering av mcnE
mcnEint6F	5' - GGG GAC GAC CAT TTT ATT TT - 3'	sekvensering av mcnE
mcnEint1R	5' - CCC AAT CAA TAT AAT CTA GC - 3'	sekvensering av mcnE
mcnEint2R	5' - GCG CTT CAA TAA ATC TGT CT - 3'	sekvensering av mcnE
mcnEint3R	5' - GGT ATT GGC AAT CGC TTT AC - 3'	sekvensering av mcnE
mcnEint4R	5' - CCA ATC TTT GTC TAA ACA GAC A - 3'	sekvensering av mcnE
mcnEint5R	5' - AAT CAA ATC GCG CTT TAA GC - 3'	sekvensering av mcnE
mcnEint6R	5' - CCA TTG ACG TTG CCA TAC GG - 3'	sekvensering av mcnE

## 5.3 Oppskrifter på løsninger

- **Alkalin PEG**

60g PEG (Sigma)  
0,96 ml KOH (Sigma)  
39 ml vann  
Juster pH til 13,3-13,5.

- **1% Agarose**

10 g Agarose (Merck) blandes med 1x TAE opp til 1000 ml. Løsningen autoklaveres i 20 min, og oppbevares i 55 °C varmeskap.

- **0,5M EDTA**

93,05g EDTA (Sigma) løses i 400ml dH<sub>2</sub>O, pH justeres til pH8 med 5M NaOH (Sigma). Volum justeres med vann. Løsningen autoklaveres i 20 min, og oppbevares ved romtemperatur.

- **50X TAE buffer**

242 g Tris-base (VWR)  
57,1 ml Iseddik (Merck)  
100 ml 0,5 M EDTA (Sigma).  
dH<sub>2</sub>O opp til 1000 ml, løses med magnetstav og oppbevares ved romtemperatur.

- **1 X TAE buffer**

20 ml 50 x TAE buffer per 1000 ml dH<sub>2</sub>O, oppbevares ved romtemperatur.

- **5M NaCl**

292.2 g NaCl (Merck) løses i 800ml vann, volumet justeres til 1 liter med H<sub>2</sub>O.  
Løsningen autoklaveres og lagres ved romtemperatur.

- **1M MgCl<sub>2</sub>**

203.31 g MgCl<sub>2</sub> (Merck) løses i 800 ml H<sub>2</sub>O, volumet justeres til 1 liter med H<sub>2</sub>O.  
Løsningen autoklaveres og lagres ved romtemperatur.

- **TB transformasjonsbuffer**

3,5 g MnCl<sub>2</sub> (Merck)  
1,1 g CaCl<sub>2</sub>·2H<sub>2</sub>O (Merck)  
9,3 g KCl (Sigma)  
10 ml 0,5M PIPES (Sigma)  
løses med 450 ml dH<sub>2</sub>O, pH justeres til 6,7, og volumet justeres til 500 ml. Løsningen sterilfiltreres og oppbevares ved -20 °C

- **SOB-medium**

20 g bacto-trypton (Merck)  
5 g gjærekstrakt (Merck)  
0,5 g NaCl (Merck)  
Løses med 950 ml dH<sub>2</sub>O, tilsett 10 ml 250mM KCl (Sigma), pH justeres til 7 med 5M NaOH (Sigma). Løsningen autoklaveres, og tilsettes 5 ml steril 2M MgCl<sub>2</sub> (Merck).

- **LB-skåler med antibiotika, IPTG og X-Gal**

- **LB-Agar**

25g LB-medium (Duchefa Biochemie)  
15 g Agar (Merck)  
dH<sub>2</sub>O opp til 1000ml

tilsatt: 4ml X-Gal (40 µg/ml)  
500ml ampicillin (Duchefa)  
30mg/ml kloramfenikol (Duchefa)  
500 µl IPTG (0.2mM) (Fermentas)

- **Ampicillin stock (25 mg/ml)**  
250 mg ampicillin (Duchefa) løses i 10 ml dH<sub>2</sub>O. Sterilfiltreres og lagres ved – 20 °C.
- **Kloramfenikol stock (30 µg/ml)**  
500 mg kloramfenikol (Duchefa) løses i 10 ml dH<sub>2</sub>O. Løsningen sterilfiltreres og lagres ved – 20 °C.
- **X-gal**  
200 mg X-gal (Fermentas) løses i 10ml dimetylformamid (Merck)  
Oppbevares ved – 20 °C i mørke eppendorfrør.
- **IPTG**  
1,2 g IPTG (Fermentas) løst i vann til total volum 50ml.  
Løsningen sterilfiltreres og lagres ved – 20 °C.
- **LB-medium**  
12,5 g LB Broth (Duchefa Biochemie) løses i 0,5 liter dH<sub>2</sub>O.  
Løsningen ble autoklavert og tilsatt ønsket antibiotikum.
- **NZYM-medium**  
22,58 g NZCYM Broth EZ-mix (Sigma) løses i 1 liter dH<sub>2</sub>O.  
Løsningen ble autoklavert og tilsatt ønsket antibiotikum.
- **1x prøve buffer (for SDS-PAGE, elektroforese av proteiner)**  
250 µl 4x XT Sample Buffer (BioRad)  
600 µl H<sub>2</sub>O  
50 µl DTT (20x Reducing agent) (BioRad)  
100 µl 1M MgCl<sub>2</sub> (Sigma)
- **1M DTT**  
3,085g 1,4-dithio-DL-threitol (Sigma) løses i 20 ml 10 mM natrium acetate (pH 5.2) (VWR).  
Løsningen sterilfiltreres og lagres ved – 20 °C.
- **Coomassie fargeløsning**  
Coomassie Blue R-250 1,0 g (Bio-Rad)  
Methanol (VWR) 450 ml  
Konsentrert eddiksyre (Merck) 100 ml  
dH<sub>2</sub>O 450 ml
- **Avfargingsløsning**  
450 ml Metanol (VWR)  
100 ml Konsentrert eddiksyre (Merck)  
dH<sub>2</sub>O opp til 450 ml
- **0,1M Fosfatbuffer**  
93,2 ml Na<sub>2</sub>HPO<sub>4</sub> (Quiagen)  
6,8 ml NaH<sub>2</sub>PO<sub>4</sub> (Quiagen)  
pH justeres til 8
- **Lysisbuffer**  
50mM NaH<sub>2</sub>PO<sub>4</sub> (Quiagen)  
300mM NaCl (Merck)  
10mM imidazol (Quiagen)
- **Vaskebuffer**  
50mM NaH<sub>2</sub>PO<sub>4</sub> (Quiagen)  
300mM NaCl (Merck)  
20mM imidazol (Quiagen)
- **Elueringsbuffer**  
50mM NaH<sub>2</sub>PO<sub>4</sub> (Quiagen)  
300mM NaCl (Merck)  
300mM imidazol (Quiagen)
- **1M CaCl**  
Løs 219.08 g CaCl (Merck) i 800 ml H<sub>2</sub>O, juster volumet til 1 liter med H<sub>2</sub>O.  
Løsningen ble autoklavert i 20 min og oppbevart ved romtemperatur.